254084

JPRS-CST-86-027

10 JULY 1986

China Report

SCIENCE AND TECHNOLOGY

FBIS FOREIGN BROADCAST INFORMATION SERVICE

3
140
A07

NOTE

JPRS publications contain information primarily from foreign newspapers, periodicals and books, but also from news agency transmissions and broadcasts. Materials from foreign-language sources are translated; those from English-language sources are transcribed or reprinted, with the original phrasing and other characteristics retained.

Headlines, editorial reports, and material enclosed in brackets [] are supplied by JPRS. Processing indicators such as [Text] or [Excerpt] in the first line of each item, or following the last line of a brief, indicate how the original information was processed. Where no processing indicator is given, the information was summarized or extracted.

Unfamiliar names rendered phonetically or transliterated are enclosed in parentheses. Words or names preceded by a question mark and enclosed in parentheses were not clear in the original but have been supplied as appropriate in context. Other unattributed parenthetical notes within the body of an item originate with the source. Times within items are as given by source.

The contents of this publication in no way represent the policies, views or attitudes of the U.S. Government.

PROCUREMENT OF PUBLICATIONS

JPRS publications may be ordered from the National Technical Information Service, Springfield, Virginia 22161. In ordering, it is recommended that the JPRS number, title, date and author, if applicable, of publication be cited.

Current JPRS publications are announced in Government Reports Announcements issued semi-monthly by the National Technical Information Service, and are listed in the Monthly Catalog of U.S. Government Publications issued by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

Correspondence pertaining to matters other than procurement may be addressed to Joint Publications Research Service, 1000 North Glebe Road, Arlington, Virginia 22201.

JPRS-CST-86-027

10 JULY 1986

CHINA REPORT

SCIENCE AND TECHNOLOGY

CONTENTS

NATIONAL DEVELOPMENTS

/9986

NATIONAL DEVELOPMENTS

SUCCESSES, FAILURES IN CHINA'S ROCKET PROGRAM REVIEWED

Hong Kong TA-KUNG-PAO in Chinese 8 May 86 p 3

[Article in column "Canadian Airmail Dispatch" column by Fan Shihe [5400 1102 0678]: "Successes and Failures in China's Test-Launching of Rockets"]

[Text] On 30 April, the Imperial Bank of Canada held in Vancouver a conference on China's Seventh Five-Year Plan. Speakers at the conference included Fan Shihe, chief engineer of China's Ministry of Astronautics Industry. He gave a detailed briefing on recent developments in China's astronautics industry, which was rich in content and some of which had not been published before. Exerpts follow:

I am honored to be able to be here today to give a briefing on the developments in China's spaceflight technology and on the key points of its Seventh Five-Year Plan. I want to first give a general picture of developments in China's spaceflight technology.

General Picture of Developments in Spaceflight Technology

Based on the state's plan for development, at the end of the fifties China began the work of opening up and developing spaceflight technology. The corresponding research institutes, designing departments, experimental centers, factories, and launch sites were gradually set up. In the early stage, we developed sounding rockets and biological rockets. These rockets were propelled by liquid rocket engines or solid boosters.

After more than 10 years of hard work, in 1970 we successfully launched our first artificial satellite, which weighed 173 kilograms and was named the "East Is Red." The carrier used for this launch, named the "Long March No 1," was a three-stage rocket. The first and second stages had liquid propellant engines. The third stage was a solid rocket engine, for spin stabilization. The second "East Is Red" satellite was successfully launched in 1971. It was launched twice, both times successfully.

In 1974, the "Long March No 2" carrier was successfully developed. This was a two-stage liquid rocket. Its length was 31.7 meters, its maximum diameter 3.35 meters, its takeoff weight 191 tons, its low orbit carrying capacity 2 tons, and its orbital inclination 63 degrees. In November 1974, its first

test launch was a failure. Data obtained from its flight confirmed that a wire on the rate gyroscope had broken. For this reason, in later design and production we used reliable circuitry and quality control. From 1975 to the present, it has made a total of seven flights, all of them successful. During this time we have also obtained successes in satellite recovery technology. Therefore, we can see that the launch reliability of the Long March No 2 carrier rocket is very high.

The Technology of Launching Synchronous Satellites Has Already Been Mastered

In the seventies China developed a two-stage carrier rocket named "Storm No 1." It looked very much like the Long March No 1, but its systems were different from those of Long March No 2 to varying degrees. Also, "Storm No 1" did not have the good performance and reliability of Long March No 2, and therefore after 1981 its use was discontinued.

"Long March No 3" is China's new-type carrier rocket. It is a three-stage liquid rocket. The first and second stages were developed on the basis of the "Long March No 2." Its third stage has liquid hydrogen and liquid oxygen engines, and it possesses the capability of being able to be switched on twice. The third stage's maximum diameter is 2.25 meters, and the diameter of its cowl is 2.6 meters. The length of Long March No 3 is 43.25 meters, and its takeoff weight is 202 tons. When inclined at 31.1 degrees, it has a carrying capacity in synchronous transfer orbit of 1,400 kilograms.

Long March No 3 was first launched on 29 January 1984. Because of a breakdown in third-stage engines, this experimental launch was not completely successful. After the first launch, we quickly found the cause of the breakdown from the telemetric data obtained during it. We used 70 days' time to modify the hydrogen rocket engine. On 8 April 1984 there was a second launch, and satisfactory results were obtained. The rocket was then used for experiments on putting a communications satellite into synchronous transfer orbit. On 16 April of the same year, the satellite was accurately positioned in space above the equator at 125 degrees east longitude. This satellite is still working today.

Following this, on 1 February 1986 a practical communications and broadcasting satellite was successfully put into synchronous transfer orbit, and on 9 February it was accurately positioned in space above the equator at 103 degrees longitude. The successful launching of "Long March No 3" proves that its design plan is correct and that it is of reliable quality. The highly difficult technology of launching a synchronous satellite has already been mastered by our country's experts and engineers.

Launching for Foreign Countries Welcomed

Up to now, our country has used its own carrier rockets at its own launch bases to launch 18 types of satellites developed by itself (including scientific sounding satellites, technical experiment satellites, and communication and broadcasting satellites). We will hereafter launch in succession our country's practical synchronous communications and broadcasting satellites, weather satellites, and scientific experiment satellites.

Next I will talk about the key points in the development of spaceflight technology in the Seventh Five-Year Plan period.

Based on our country's practical needs, we will in succession put the "Long March No 3" and "Long March No 4" into service for our country's consumers. At the same time, based on our country's open-door policy, we have decided to provide "Long March No 2" and "Long March No 3" for use in international commercial launches. We welcome foreign consumers who use satellites for peaceful purposes to entrust us with the carrier launch. "Long March No 2" has a very good success rate. We intend to further improve the performance of Long March No 2. There is a high degree of consistency in the design, technology, quality control, as well as systems engineering management in our country's Long March series.

To further satisfy the needs of consumers, our country will further improve the "Long March No 3" carrier vehicle, so that it has a carrying capacity of more than two tons when in synchronous transfer orbit. We will modify its control system so that it has the capability to adapt to orbit changes. At the same time, to launch heavier synchronous satellites, we will add a booster to the I class, so that the Long March carrier vehicle becomes an integrated system.

The satellite is also one of our country's key development projects. in the Seventh Five-Year Plan period, first we will develop communications and broadcasting satellites with large capacities. The satellites will have a triaxial stabilizing, double-unit element unified drive system program. Second, synchronous weather satellites and earth resources satellites can be used for weather forecasting, natural resources prospecting, and soil data surveys as well as environmental monitoring. Third, to meet the needs of consumers, we will develop a large number of ground stations with small diameter antennas in order to develop our country's communications and broadcasting undertakings.

The above are the main points in our country's development of spaceflight technology in the Seventh Five-Year Plan period.

Providing Good Service

To put into effect the open-door policy and carry out international cooperation is our country's national policy. On the basis of acting independently and keeping the initiative in our own hands, being self-reliant, equality and mutual benefit, and mutual trust, we will carry out exchanges and cooperate with spaceflight circles of various countries in the world. Provided the two sides have the need and the possibility, we will introduce technologies, introduce equipment, make extensive use of the strong points of various countries, and strive to raise the scientific and technical level, the administrative and management level, of our country's spaceflights. In the past several years we have signed spaceflight scientific and technical agreements with four countries in Western Europe--West Germany, France, Britain, and Italy--and we are now holding cooperation talks or signing agreements with some other countries on the European, Asian, and American continents.

Canada is a developed country, and its spaceflight technology is advanced. We hope China and Canada will henceforth be able to initiate various forms of cooperation. Our country has already decided to give play to its own superiority in spaceflight technology by putting its carrier rockets on the international market and by contracting to provide services in the development and launching of carrier rockets for the satellites of various countries. At the same time, we are willing to initiate various forms of cooperation with foreign countries in launch, measurement and control technologies.

The reasons that we are carrying out international cooperation and providing launching services are:

First, our preferential prices are lower than the preferential prices on the international market.

Second, at each stage of the launching service, we will provide good service to the greatest degree possible.

Third, we provide insurance. The China People's Insurance Corporation is willing to provide preferential insurance for any project proposed by a customer.

9727
CSO: 4008/63

NATIONAL DEVELOPMENTS

XI'AN JIAOTONG UNIVERSITY CALLED MAJOR SCIENCE BASE

Beijing RENMIN RIBAO in Chinese 7 Apr 86 p 1

[Article: "Xi'an Jiaotong University Has Taken Root in the Northwest, in 30 Years Becoming Its Major Educational and Scientific Research Base; Over 31,000 Graduates Have Been Trained and 570-plus Scientific Research Gains Achieved in the Past 15 Years"]

[Text] Correspondent Jiang Hanzhen [5592 5496 4631] reports that Xi'an Jiaotong University, which was begun in Shanghai in 1956 and then relocated inland, has now passed its 30th year. It has already taken root in the Great Northwest and become one of the major northern bases for training senior technologists and fostering scientific research.

To meet the needs of socialist economic construction and development in China, and to change the past irrational distribution of institutions of higher education in China, in 1956 the State Council decided to move Jiaotong University to Xi'an to help with socialist economic construction in the interior. At that time Premier Zhou Enlai personally took charge of the task, and he told the teachers and students that "The Chinese revolution originated in the Southeast and came to fruition in the Northwest. Construction in the Northwest must be aided by industrially and technologically advanced coastal areas." Premier Zhou hoped that the intellectuals at Jiaotong University would pluck up the courage "to ride a while on the back of a tiger," brave the storm, and face the world to move the great tree, which had grown for 60 years in Shanghai, to the Northwest, there to root it deeply in the fertile Northwestern soil.

Fired by Premier Zhou's enthusiasm, several established professors quickly straightened out property that had to remain behind in Shanghai, abandoned the good life and working conditions of Shanghai, and packed their wives and children off to Xi'an. Thirty years ago living conditions in Xi'an were rather difficult and food and necessities all had to be brought in from Shanghai. They uttered not a word of complaint, but wholeheartedly carried out their teaching duties and ensured that there was no drop in educational standards after the school was relocated to Xi'an. By 1958 most of the faculty had arrived in Xi'an, including 43 professors and 22 assistant professors. Thirty years ago, Professor Zhong Zhaolin [6988 0340 3829], now 86, was an activist in the campus relocation. He is now in a Shanghai

hospital for rest and recuperation. Professor Zhong is constantly mindful that he is a "Northwesterner" and he communicates periodically with the school. When he is well again he desires to return to the Northwest to work. Former Dean Peng Kang [1756 1660], who led the campus relocation at that time, and some well-known former professors devoted the second half of their lives to the Great Northwest and have long graced Xi'an's cemetary for revolutionary martyrs. The young students at Jiaotong University were even more prone to think of "transfer to the Great Northwest" as everybody's highest ideal: when the orders came there were no foot-draggers. Thirty years have passed, and second generation students at Jiaotong University have already married and embarked on their careers, while the third generation has become a true "little Shaanxi."

While assisting the Great Northwest, Jiaotong University has also achieved its own growth. At the time of relocation Xi'an Jiaotong University was a college of engineering, featuring 4 departments and 15 specialties. Now it has developed into a combined science and engineering-liberal arts university, emphasizing engineering, with 18 departments and 46 specialties. From 700-plus teachers at the time of relocation, the school has grown to more than 1,800 teachers, including 600-plus professors and assistant professors. In the 30 years since relocation the university has trained over 31,000 graduates--4 times as many as in the 53 years prior to the founding of the PRC--and one-third of these have been assigned in the Great Northwest. During the Sixth Five-year Plan, universities under all branches of the central government trained 12,000 college students for Shaanxi, and 20 percent of these were trained at Xi'an Jiaotong University. Now the technological backbone of some large northwestern enterprises, such as Lanzhou Refinery, Lanzhou Petroleum Machinery Plant, and Xi'an Electric Power Machinery Company, is composed of students trained at post-relocation Jiaotong University.

In the past 15 years the school has garnered 577 achievements in scientific research, of which 88 reach or approach international research standards, 2 earned National Natural Science Awards, and 10 won National Invention Awards. Economic benefits from existing scientific research achievements approach 500 million yuan, of which one-fifth goes to relevant factories, mines, and other enterprises in Shaanxi.

The educational base in the Great Northwest is rather weak, and after Xi'an Jiaotong University moved there it played an exemplary role in developing northwestern higher education facilities. It signed a long-term cooperation agreement with schools of higher education in the minority nationality region, helping those schools to improve their standards rapidly. In the past few years, it has trained 400-plus English teachers for institutions of higher learning in the northwestern region. Right now there are, altogether, 43 doctorates offered at the 40-plus institution of higher education in Shaanxi, and Xi'an Jiaotong University accounts for 19 of them. This school has become Shaanxi's primary forum for launching academic exchanges with the outside.

In the past 30 years the people of Shaanxi have shown zealous concern for Xi'an Jiaotong University, and in difficult times millet from Yan'an and vermicelli and white bread from the central Shaanxi plain have fed its teachers and students. Many times the people's government of Shaanxi has

organized efforts to transport prawns and fresh fish back from coastal areas to give comrades from the South a taste of seafood. As for the school's capital construction, Shaanxi spared no effort and completed the task in short order.

When Jiaotong University moved west, some comrades worried about whether relocation policy formulated by the central authorities was correct. At the time, comrade Chen Yi [7115 3015], mayor of Shanghai, said: "The verdict on relocating Jiaotong University will come after 10 years." Thirty years of experience amply illustrates that the relocation of Xi'an Jiaotong University has been a successful model for China in readjusting the strategic distribution of institutions of higher learning. The construction and growth of Xi'an Jiaotong University has far-ranging significance for development in the Great Northwest.

12510
CSO: 4008/2100

CHINA UNIVERSITY OF S&T REPORTS 60 PROMOTIONS

Beijing RENMIN RIBAO (Overseas Edition) in Chinese 13 Mar 86 p 3

[Article: "At China University of Science and Technology Commendation Is Based on Talent Alone: 60 Younger Teachers Are Promoted to Professor and Assistant Professor"]

[Text] According to a report in ZHONGGUO QINGNIAN BAO, "China University of Science and Technology is well-run and boasts plenty of young talent." In the wake of this praise, proferred by comrade Deng Xiaoping in 1984, China University of Science and Technology further eradicated the outworn concept of rank ordering based on seniority, persisted in considering academic standards, and emphasized practical contributions. Recently it also boldly promoted 60 younger teachers to the ranks of professor and assistant professor, a move which evoked strong reverberations among instructors and students at the school.

Among these 60 individuals, 6 instructors under the age of 45 were promoted to professor and 47 were promoted to assistant professor, while an additional 7 younger teachers under the age of 40 were promoted to assistant professor. Five of this latter group hold doctorates earned in Chinese universities. They have studied diligently, are imbued with the spirit of initiative, are constantly updating their knowledge, and have become the school's young academic leaders.

Chen Lin [7115 7207], who skipped a rank and was promoted to professor, is only 40. Altogether, since 1976 Chen has published over 20 insightful articles. In 1980, when he was a visiting student in the United States, his original concept of visual topology and functional layers was appraised very highly by associates from many nations in the fields of psychology, neuroscience, and computer visualization. He has been called the world's "first person to combine visual perception with topology."

12510
CSO: 4008/2100

NATIONAL DEVELOPMENTS

LU JIAXI DISCUSSES NEED FOR S&T COORDINATION

Beijing GUANGMING RIBAO in Chinese 2 Apr 86 p 2

[Article by Li Shuxi [2621 2885 0823] and Zhai Huisheng [5049 1920 3932]: "Lu Jiaxi [4151 0857 6932], Member of the CPPCC Standing Committee Discusses Building Up the Main Force of Science and Technology Through Coordination"]

[Text] When reporters met with Lu Jiaxi [4151 0857 6932], member of the CPPCC standing committee and president of the Chinese Academy of Sciences, he immediately began to discuss reform in the system of scientific and technological organization and problems in coordination.

He said that when comrade Yao Bang talked about thousands and thousands of soldiers and horses climbing the mountain to pick peaches, he evoked a vivid image. However, there cannot be only one road up the mountain. If many thousands of soldiers and horses are crowded onto a single road many will be trampled; if there is only one kind of peaches to pick the great army will be unable to exercise its function. Supposing that people have not carefully planted a variety of peaches in advance, this army of men and mounts cannot make its best effort. Therefore, we must recognize "diversity of form on all sides."

Lu Jiaxi has visited abroad on several occasions recently. What he saw and heard there made it even clearer to him that scientific and technological development strategy, and reform in the system of scientific and technological organization, must be geared to the needs of the four modernizations, the world at large, and future ideology. He said that while abroad he heard people saying that there must be the three "co's": coordination, cooperation, and collaboration. Coordination is comprehensive: rational vertical and horizontal relationships are established and scientific and technological administration must weave these into a network. There are many "scientific" organizations and units in China, and particularly those at the first level of central government should divide the work and cooperate with one another. Right now there is too much duplicate labor: one is doing the same thing another is doing, and sometimes the two are kept secret from each other. There is no surfeit of talent and funding in the first place, so duplication is a great waste, bringing half the result with twice the effort. We should consider the situation as a whole and then make certain choices. There are also many problems in cooperation: there are frequent disputes over trifles,

to the extent that good works are often left in a terrible mess. Herein lie problems in administrative systems and in ideological and political training.

Lu Jiaxi extracted from his bundle of papers the report made by Premier Zhao concerning the Seventh Five-year Plan, and told reporters that the system of scientific and technological organization must be reformed to serve economic construction. First of all we must concentrate on resolving major current research problems in production and construction. Previously we have devoted insufficient attention to this area, and hereafter we must greatly enhance our efforts. But even this is not enough: science and technology must precede production and construction. We must consider the economic construction and services of tomorrow, and we must think about laying the scientific foundation and providing technological guarantees for sustained, stable economic growth through the next 10 years and the 1st 50 years of the 21st century. Deng Xiaoping has pointed out that education must be geared to modernization, to the world at large, and to the future. This demand also applies absolutely to scientific research. In achieving this goal the crucial issue is still one of talent: how to stimulate initiative among scientific and technical personnel, and how to build up the main talent force. This is strategic, for it takes 100 years to build a generation of good men! The main talent issue is to assign people according to their aptitude. People have different talents: some people are clever (theoretical thinking comes easily) and some are deft (testing skills are strong); some people are good at "quick battles" (resolving short-term, ordinary, straightforward problems) and some can fight a "protracted war" (systematically resolving major problems). So long as it benefits the country and the people, we should encourage them to go all out in their work. Of course we must take all factors into consideration in overall planning and make suitable arrangements so that the appropriate quantity and quality of talent is present in all areas of basic research, applied research, and development. We must be adept at nurturing, discovering, and organizing scientific research forces, and we must take particular care to give those persons engaged in primary exploratory research and education the opportunity to participate in high-level research and teaching activities. We must be capable of interchange and permit people to change their occupations. At home and abroad there are a number of well-known scientists and scholars who have shifted occupations, and there is nothing wrong with this. Our current arbitrary uniformity and overcentralization disadvantage people in the exercise of subjective initiative.

Finally, Lu Jiaxi said that talent depends a person's own diligence, as well as on training and policy; diligence creates an opportunity for him to display his talent. "I have had a long career in teaching. I often say that if a teacher cannot train a number of students to be better than himself then he should be deposed." At that point, Lu Jiaxi laughed and said: "I have many students who have made outstanding contributions in their respective fields of scientific research and a few students who are now in Beijing attending the national CPPCC and NPC meetings, among them Xiamen University President Tian Zhaowu [3944 2507 2976]. There is a group of students who have surpassed me, so it looks like this particular teacher will not be deposed."

12510
CSO: 4008/2100

NATIONAL DEVELOPMENTS

S&T SERVICE TO FOUR MODERNIZATIONS DISCUSSED

Beijing GUANGMING RIBAO in Chinese 2 Apr 86 p 2

[Article by Wu Qinying [2936 0530 5391]: "How the Ranks of Science and Technology Can Serve the Four Modernizations Is Discussed by Wang Daheng [3769 1129 3801], NPC Deputy and Chinese Academy of Sciences Academic Department Member"]

[Text] Professor Wang Daheng [3769 1129 3801] is currently a deputy director of the Second Technical Sciences Department of the Chinese Academy of Sciences, and the honorary director of the Changchun Optics and Fine Mechanics Institute. In February and March of 1986, through an investigation of the implementation of scientific funding in the Chinese Academy of Sciences, Wang became deeply convinced that the issue of scientific and technical personnel serving the four modernizations is extremely important. He discussed his views with reporters:

I. It Is Well Worthwhile To Gear Scientific Research Toward Production

Professor Wang Daheng said an example of this kind is that the Military Academy of Medical Science used genetic engineering techniques to fabricate, at an advanced international level, an engineered vaccine for swine dysentary. They mixed this vaccine into fodder and fed it to sows, and the surviving swine possessed immunocompetence against yellow and white dysentary. According to calculations, if this success was extended throughout the country 1.5 billion yuan would be gained per year. Professor Wang said that many scientific and technical personnel are now engaged, under crude and simple conditions, in research on the key problems that urgently need to be solved, and number of high-level achievements have been made. There is, for example, Qinghua University's "new welding arc control technique," which has brought welding automation and achieved major economic results. Another example is the "Land Utilization Chart" drawn up by tens of units and hundreds of scientific workers nationwide, organized by the Chinese Academy of Sciences. It has provided an important basis for district planning and land development in China.

II. Qualified Scientists and Technicians Must Suit the Needs of Scientific Socialization

Wang Daheng said that nowadays a major scientific research project more often than not requires mutual cooperation between tens of units and as many as a hundred experts engaged in mutual interdepartmental and inter-unit cooperation. This requires scientific and technical personnel both to be strictly imbued with scientific character and to have the capacity for horizontal alliances featuring mutual interchange and cooperation. Wang Daheng is very cognizant of the qualities of scientific and technical personnel, and he says that if one cannot easily cooperate with others, then one cannot suit the needs of scientific socialization. We need experts who can take charge, and, what's more, we need experts who are well-connected in society as a whole and in scientific and technological relationships.

III. Give Gifted Members of the Scientific Community a Role To Play

Wang Daheng emphasized that there are a number of Academic Department members in the Chinese Academy of Sciences who are rare talents in the scientific community. We must give these people a role in important issues of scientific and technological development and economic construction. For example, important scientific projects and major construction projects must all be subjected to the arguments of experts. Through these people we can organize a multitude of forces and link up with talented people in various areas. The Academic Department members enjoy high prestige, command universal respect, have a wealth of experience, and possess profound knowledge. In addition, a large number of them have left their administrative posts. If we can give them a greater role to play we can reduce errors in construction and promote smooth construction in the four modernizations.

12510
CSO:4008/2100

HORIZONTAL RESEARCH, PRODUCTION ALLIANCES URGED

Beijing GUANGMING RIBAO in Chinese 30 Mar 86 p 1

[Commentary by Yu Weidong [0060 4850 2767] under the rubric "Weekly Commentary": "Develop Horizontal Research and Production Alliances"]

[Text] Modernized production is socialized production built upon a foundation of advanced science and technology. It is characterized by daily, intimate cooperation and integration between research and production. Cooperation between research and production is not something we have just acquired: since the founding of the PRC, the 156 projects in the 1950's and the "2 Missles, 1 Satellite Project" of the 1960's were successful examples of large-scale cooperation between research and production. However, this cooperation was brought about by administrative means; it might be called a tree-like construction. In the 1980's cooperation between research and production in China has taken on the new form of horizontal alliance. The horizontal alliance is characterized by unchanging ownership and subordinate relationships among member units, the use of economic leverage, and conformity to the principles of voluntary participation, equality, and mutual benefit. It considers the development of new products, techniques, and technology--or technological reform--to be its objective. The alliance is a new coalition connected by economic ties; one could say it is a network-like construction. According to incomplete statistics, as of the end of 1985 there were over 10,000 allied research and production organizations of all kinds nationwide, of which 2,000-plus were in Shanghai and 1,200-plus were in Hebei. Horizontal research and production alliances occur in the following forms:

As to distinguishing the degree of intimacy in these alliances, we can divide them into two types: the loose alliance and the substantive alliance. The loose alliance binds a research unit and a production unit through a project contract and they jointly develop new products or new techniques. When the project contract is completed the alliance will have completed its mission as well. Consequenlty, this is a temporary, low-level form of alliance. In the substantive alliance, the two parties or the various parties in the alliance take part in joint capital or collaboration and organize themselves into an economic entity that operates jointly and divides profits. This form of horizontal alliance is also called an economic association. For example, Dalian Economic Association for Producing Textile Printing and Dyeing is made up of 16 units, including schools of higher education, research units, and enterprises. It achieves the full course of complementary research and

production: sample analysis, synthesis of raw materials, manufacturing of auxiliary textile chemicals, reeling and spinning of raw silk from cocoons, weaving, and dyeing. In 2 years time this economic association has come out with 90 new textile products.

As for distinguishing the substance of the alliances, we can divide them into the following types: alliances for industrial development, enterprise transformation, new product manufacturing, digestion of imported technology, new technological development, achievement transfer, project contracting, and multipurpose use of natural resources.

As to distinguishing the scope of allied member units, we can differentiate between two types: domestic alliances and international alliances. Domestic alliances can be further divided into the urban form (including alliances between research and production units within one city, and alliances that cut across regions, provinces, cities, and departments) and the integrated urban-rural form. In the wake of China's open door policy there has also been new growth in international research and production alliances. For example, Shanghai Materia Medica Institute of the Chinese Academy of Sciences and Japan's Comprehensive Medicinal Plant Testing Association allied to develop the new antineoplastic drug "AT-2153," and Shanghai Silicate Institute and Shanghai Xinyu Power Plant allied with the American Energy Conversion Device Company to develop the noncrystalline silicone solar cell. These are both examples of development through international alliances.

As far as distinguishing the allied member units is concerned, we can divide these into two-unit types (a research unit and a production unit or an institution of higher education and a production unit) and multi-unit types (research unit, institution of higher education, and production unit).

In the past few years horizontal alliances between research and production have expanded rapidly. They are just now unfolding, and are demonstrating great vitality. Their primary functions are as follows:

1. They have promoted technological progress in enterprises, particularly in medium-sized and small enterprises and township enterprises, they have stimulated a number of small factories, and they have brought clear economic results.

2. They have spurred research units and institutions of higher education to be more attuned to economic construction. Moreover, these units have increased receipts, enhanced self-development capabilities, improved research and education standards, and been advantageous in training new talented personnel who integrate theory with practice.

3. They have advanced intelligent and rational circulation, and have been helpful in allowing existing talent to exercise its role. Horizontal alliances between research and production do not change the subordinate relationships of personnel. They have enabled a great many scientists and technicians to go beyond the tall buildings and spacious courtyards of the research institutes and schools of higher learning, enter the front lines of production, and display their skills in medium-sized and small enterprises and

township enterprises. They go gladly, of their own free will, and this is something that was impossible to attain in the past under purely administrative means.

We must further develop alliances between research and production, and we must also resolve the following issues:

1. We must establish and perfect pertinant laws and regulations, safeguard the legal rights and interests of the allied parties, and ensure healthy development in research and production alliances. Right now, some scientific and technical departments report that the contracts signed frequently will not hold up; by contrast, production units report that some scientific and technical achievements are not ready yet, thus causing the production units to sustain losses.

2. We must resolve policy and organization problems. How can we give the appropriate support to public finanaces, tax revenues, and banking? We still lack a comprehensive set of policies. For example, intermediate testing is the link in transforming research achievements into industrialized production, but under the existing system of organization research units and enterprises alike find it a difficult problem to tackle. We must reform the system of organization now in effect, clear out funding channels, and create appropriate social conditions for the growth of research and production alliances.

3. We must enhance theoretical research. Horizontal alliances between research and production are based on commercialization of technical research achievements. However, in China we still do very little theoretical research on commercialization of scientific research gains. We understand very little about the specific properties of technical merchandise, or about the law of value as it concerns these goods. This leads to price confusion and very uneven remuneration in the technological marketplace. Many new problems arise out of this, and we need to have a proper theory to guide us in resolving them.

The growth of horizontal alliances between research and production is required for modernized production, for a socialist commodity economy, and for rapid forward development in research facilities. The extensive development of horizontal links is in itself a challenge to the original disjointed situation; it is a powerful assault on the old system of organization. In a sense one could say that the process of converting the old system to the new system of organization is precisely this process of eradicating the disjointed situation and extensively developing horizontal economic links, including horizontal links between research and production. Therefore, the development of horizontal alliances between research and production is extremely important for promoting economic restructuring in China's urban and rural areas, and also for promoting reforms in the system of scientific and technological organization and the system of education. We must enthusiastically foster and support this effort and point it toward a new stage of extensive and healthy development.

12510
CSO: 4008/2100

NATIONAL DEVELOPMENTS

JIANGXI RURAL TECHNOLOGY MARKET RESULTS REPORTED

Beijing NONGMIN RIBAO in Chinese 18 Mar 86 p 1

[Article: "Jiangxi Is Quickly Guiding Science and Technology Toward the Countryside, Obtaining Outstanding Results from the Rural Technology Market; 62.57 Million Yuan Worth of Composite Social and Economic Benefits Were Created in 1985, About 20 times the Amount of Scientific and Technological Investment"]

[Text] In order to guide science and technology toward the countryside, Jiangxi's Science Commission began in 1984 to establish a provincewide rural technology market. At present it has grown to encompass 98 counties (or administrative areas), accounting for 88 percent of all Jiangxi counties (or administrative areas). In 1985, 62.57 million yuan worth of composite social and economic benefits were created, about 20 times the amount of scientific and technological investment.

All over Jiangxi the core of the technology market is the county, which connects above to the province and region and below to townships and villages. Each market utilizes a combination of administrative means and economic leverage to enlist the services of talented people from social and scientific research units and institutions of higher education. They collect technology and vitalize rural economic services by organizing technological service teams and technological training, and by establishing economic associations between technology and production. For example, Longnan County Science and Technology Market called together over 60 scientists, technicians, and skillful craftsmen and organized 7 technological service teams, each with a different specialty, to go out into the countryside. Of these, the Technological Service Team To Transform Low-yield Orchards had 33 members, and they signed service contracts on nearly 30,000 fruit trees grown by 332 fruit-farming households. That year, through technological reforms--pruning, cultivation, fertilization, insect eradication, and disease treatment--the per-mu yield in Chinese chestnut orchards increased 50 percent over the previous year, and the per-mu yield in citrus orchards rose 130 percent. In Ganzhou Prefecture the technology market conducted a technical training class on pond milkfish breeding, and more than 3,200 people attended the class. After the students left the class they set off a surge of enthusiasm for raising pond milkfish. In 1985, 2.4 million-plus jin of milkfish were harvested throughout the prefecture. A collective of only four or five people set up shop as the

Ningdu County Chalk Factory and suggested the trial manufacture of dust-free chalk. However, they lacked the technology, funding, and equipment for the job. Ningdu County Technology Market sent out technical personnel, invested 13,000 yuan, and entered into joint operation with this plant to develop the product. Now this plant has grown into a new-style small enterprise employing 47 people. Its product sells well in Ruijin, Yudu, and other counties, and has also broken into the market in Nanchang, Shanghai, and other regions. In 9 months of 1985, this plant reached a production value of 95,000 yuan and achieved a net profit of 25,000 yuan. According to statistics, in 1985 Jiangxi's rural technology market launched 40,690 technological service projects (or instances) of all sorts, trained more than 15,000 youth, established 112 technical economic associations of various kinds, and imported and transferred 373 technical items.

In order to enhance guidance of rural technology markets and arrange coordinated efforts, the Jiangxi provincial government has formulated provisional methods for managing technology markets, and they have also recently approved the establishment of a provincial technological market development center.

12510
CSO: 4008/2100

MORE TECHNOLOGY IMPORTS FOR MACHINE BUILDING INDUSTRY

OW291907 Beijing XINHUA in English 1757 GMT 29 May 86

[Text] Beijing, May 29 (XINHUA)--China is to import another 800 items of technology to revamp its machine building industry with foreign funds to be absorbed in the next five years.

This was disclosed by He Guangyuan, vice-minister of machine building industry, at today's national conference on machine building industry, which undertakes an important task of providing equipment for petroleum, coal, chemistry, electricity and textiles industries.

Priorities for technical import will be given to large and high-capacity equipment, such as mining, petrochemical, and energy-saving projects and those for producing export goods, the vice-minister said.

The industry will also make big efforts to attract more foreign investors, he said. It will first go into a group of 28 enterprises in cooperation with foreign partners in the form of joint venture, co-management and sole foreign investment in the next five years, which concern high technology and their products are export-oriented.

Besides, negotiations are under way on some projects between foreign businessmen and China's localities.

China's machine building industry will also make use of foreign loans in building and renovating 16 projects, including the projects of the Shanghai Machine Tool Plant and the Wuhan Heavy Machine Tool Plant.

According to the ministry, China imported 800 items of technology between 1978 and 1985, and now 47 percent of them have gone into production and 33 percent produced sample products.

By technical import and technological development, the machine building industry will stress upgrading its existing enterprises to manufacture complete sets of equipment of energy, transport, raw materials, and instruments and meters.

--In the fields of electricity generating and high-voltage transmission and transforming equipment, the industry will revamp four power centers of Beijing, Shanghai, Harbin, and Sichuan, and two transmission equipment producing centers in Xi'an and Liaoning. The industry will have an annual capacity of producing 10 million-kilowatt generating equipment by 1990.

--On the iron and steel industry, the industry will upgrade six heavy machinery plants in Shanghai, Shenyang and elsewhere in China.

--For the automotive industry, China will build some motor vehicle producing centers in Beijing, Changchun, Nanjing, Shanghai, and other areas. The annual output of motor vehicles will reach 700,000 in 1990.

--The production of equipment and machinery for open-cut coal mines, off-shore oil development, plastic processing and farming machines will also be expanded.

/6091
CSO: 4010/59

NATIONAL DEVELOPMENTS

RETURNED STUDENTS ACHIEVEMENTS, ROLE IN RESEARCH NOTED

Beijing RENMIN RIBAO (Overseas Edition) in Chinese 21 Apr 86 p 4

[Text] In order to build up a highly qualified teaching force, the China University of Science and Technology has in the past few years sent 600 people including 310 teachers to 18 countries for advanced studies. Now 183 have returned after completing their studies abroad, and they are playing a prominent role as teachers and in scientific research.

According to a KEXUEBAO report, the students sent abroad by this university generally have done fairly well. Wang Zhaozhong [3769 5128 0022], who was sent to the 11th University of Paris, France, in 1979, discovered a new phase change in the conducting material three magnetization tantalum, which is known as one of the world's 10 discoveries in that field. Wu Zhongchao [0702 1813 6389], while studying at Cambridge University, England, won third place in the 1985 treatise awards sponsored by the International Foundation for Gravitational Studies. Cao Xiren [2580 1585 0088], while studying at Harvard University in the United States, won an award for outstanding student theses in America and had 13 highly regarded academic treatises published in influential U.S. magazines. The 183 returned teachers wrote a total of more than 600 treatises while abroad, more than 400 of which were of fairly high standards and published in international publications and journals. By their performance abroad, they could be divided roughly into three groups, with 40 percent of them in group A, 53 percent in group B, and 7 percent in group C. Based on the university's requirements, they generally took some new courses to strengthen their theoretical and professional knowledge. In research work, they made full use of the advanced laboratory and technological conditions abroad. Thus they have greatly enhanced their knowledge and training in both theoretical and technological fields and laid a good foundation for work after their return.

The teachers began to return to China one after another at the end of 1980. Most of them have played a backbone role in work, and many have become academic leaders. They impart the new technology and knowledge they have acquired from abroad in class and research work to the welcome of teachers and students.

According to incomplete statistics, they have opened nearly 100 new courses and run more than 20 discussion and research classes of various types. More than 60 of the 183 returned teachers are serving as tutors for students working

toward a master's degree, and 70 percent of them are teaching while doing research work. Since their return, they have published more than 400 academic papers; obtained approval or submitted report on 87 research projects including major research projects and projects financed by scientific foundations, which account for 53 percent of the total number of similar projects undertaken by the university as a whole; and produced results on 34 research projects, 6 of which won major achievement awards. Most of them have become academic or administrative leaders; 13 have been promoted to professors; 70 have become associate professors; and 47 have become the number one men at various levels, including 31 department heads.

12802/12859
CSO: 4008/2107

CONTINUED SUPPORT OF TECHNOLOGY MARKET URGED

Beijing GUANGMING RIBAO in Chinese 7 Apr 86 p 1

[Article by staff commentator: "We Must Continue To Give Vigorous Support to the Technology Market"]

[Text] A year ago, we said that the appearance of the technology market is an objective requirement of the developing socialist commodity production in China, an inevitable product of the commercialization of science and technology. Practice in the past year has borne out that since the publication of the party Central Committee's decision on the reform of the science and technology management system, it has been the technology market that acts most promptly and produces the greatest impact. As a link between technology and economics and a bridge for scientific and technological achievements to be popularized and applied in the field of material production, the technology market has taken root firmly in China and has born rich fruits.

The situation is good as a whole. There are more than 5,000 technology trade organizations in the country, and their actually completed trade volume was 2.3 billion yuan in 1985, more than three times that in 1984. The technology trade is becoming increasingly extensive in scope, rich in content, and varied in forms. It has also been further strengthened in organization and management. Every province, municipality and autonomous region has set up its own organization dealing in technological commodities. More than 3,000 technology fairs of various forms were held in the past year. Technology trade has developed from transfers of single technological achievements and general technical services to technical training, technological job contracting, offers of technology as shares in joint ventures, joint development of new products and so forth. A large number of advanced and applicable technologies have been well received by small and medium-sized enterprises and town and township enterprises. Some items originally planned to be imported from abroad are now being obtained from the technology market through competitive bidding.

The results are obvious. The development of the technology market has promoted cooperation between scientists and producers, increased the ability of scientific research units to develop on their own, brought new vitality to the town and township enterprises and given scientists and technicians room to put their abilities and talents to use.

The technology market has unclogged the channels for scientific and technological achievements to flow into production so that they can be turned more quickly into real productive forces. The transfer rate of scientific and technological achievements in Shanghai in 1985 was double that in 1981, and more than 40 universities and colleges earned a net income of 80 million yuan last year from trade in technology. Through the technology market, a great deal of technology has found its way into the countryside, giving a very great impetus to rural economic development. Lingchuan County, Shanxi, had long been dependent on state financial subsidies, and then it became possible for the county to bring in technology, trained people and funds from the technology market. In 1985, the output value of the county's rural enterprises broke the 100 million yuan mark.

All this tells us that the development of the technology market is a major event in the history of the development of science and technology in China, and that it is playing an important role in propelling China's scientific, technological and economic growth.

The development of the technology market is a policy decision of the party Central Committee, and it is a completely new thing. However, precisely because it is still in the early period of development and lacks experience, it is unavoidable that problems of one kind or another will occur. It will be easy to go back to where we were before, but there is no hope in so doing. The correct attitude can only be to strengthen our confidence and actively support the technology market, analyzing, studying and properly guiding it, so that it will keep improving. The activities of the science and technology management system, the technological development institutions and the intermediary organizations in technological transfers, technological consultations, technological services, technological training and so forth must not be confused with the unhealthy trends in party and government organizations and among party and government cadres like using one's position and power to seek personal gain and feathering one's nest at public expense in doing business and running enterprises. At present, it seems that we still must adhere to the principle of "relaxation, flexibility, support and guidance" in developing the technological market. When we stress the need to draw distinctions in accordance with the party's policies, pay attention to legislative work, strengthen management of the technology market, attach importance to the development of complete sets of technologies, do a good job in the training of managerial personnel and so forth, our purpose is to implement this principle correctly.

The technology market is developing from small to big, from a lower to a higher level and from imperfect to more perfect. The burden is heavy and the road is long. The prospects are broad. We must still work hard.

12802/12859
CSO: 4008/2107

NATIONAL DEVELOPMENTS

SHANGHAI CAS PROMOTES SCIENTIFIC COOPERATION

Beijing GUANGMING RIBAO in Chinese 7 Apr 86 p 1

[Article by staff reporter Huang Dongyuan [7806 0392 0337]]

[Text] In the course of the reform of the science and technology management system, the research institutes under the Shanghai branch of the Chinese Academy of Sciences have paid attention to bringing their superior technology and qualified people into play and actively developing high-level horizontal scientific and technological ties and cooperation in the form of partnerships between research and production, transfer of technological achievements, technological consultations and so forth. In the past few years, they have provided a number of urgently needed technological achievements for economic development and some new materials and products. Some of the products have found their way into international markets.

The high-level horizontal relations being developed by the research institutes of the Shanghai branch of the Chinese Academy of Sciences are mainly in the following three areas:

The first is to step up the work to develop new technologies. In the past few years, micro-electronics, optical fiber communications technology, genetic engineering and new materials have developed rapidly in China. Some research institutes have actively cooperated with related Shanghai factories in solving key technological problems. They have done a great deal of research work and achieved some results in the use of integrated circuits in Chinese-made video cassette recorders, color television sets, tape recorders and so forth. The Shanghai Metallurgy Institute, Shanghai Optics and Fine Mechanics Institute, Shanghai Communications Equipment Plant and Jiaotong University have worked in concert and successfully developed the terminal equipment and photoelectric components required in optical fiber communications. And they have now developed the capacity to produce 200 sets of various types of terminal equipment and several hundred sets of various kinds of photoelectric components.

The second is to cooperate with the large key enterprises. In 1984, The Shanghai Branch of the Chinese Academy of Sciences and the Shanghai Petrochemical General Plant signed a "scientific and technological cooperation agreement" to jointly tackle key technological problems and popularize research achievements. Up to the present, three of their technological achievements have been translated into fairly good economic results. The Shanghai

Metallurgy Institute has cooperated with the Shanghai Baoshan Iron and Steel Complex in conducting research on the corrosion of metals used in the Baoshan Iron and Steel Complex's water diversion works, and has provided the scientific basis for the iron and steel complex to use the Changjiang River's water for industrial purposes. According to calculations by experts concerned, this project has saved an investment of at least 50 million yuan for the state.

The third is to serve enterprises in importing, digesting and applying technology from abroad. They have cooperated with other units and successfully developed a number of new products and materials to substitute for imported products and materials. The Shanghai Organic Chemistry Institute has successfully developed an antifreeze liquid for motor vehicles, which had to be imported in the past. It is now produced in batches by a factory in Nantong, Jiangsu.

12802/12859
CSO: 4008/2107

NATIONAL DEVELOPMENTS

YOUNGER SCIENTISTS, TECHNICIANS PLAY LARGE ROLE AS INVENTORS

Beijing BEIJING KEJIBAO in Chinese 26 Mar 86 p 1

[Article by staff reporter Shi Wenjie [0670 2429 1240]]

[Text] On the eve of the first anniversary of the promulgation of China's Patent Law, this reporter learned from the Municipal Patent Bureau that up to now more than 60 inventions have been granted patents in the Beijing area. Among the scientists and technicians who have won patent rights, the overwhelming majority are people under 50. Young and middle-aged scientists and technicians are playing a major role as patent-winning inventors in Beijing.

These young and middle-aged inventors not only aspire to invent and create things as their goals in life, but link their own inventive and creative activities closely with the destiny of the motherland, the future of socialist modernization and the interests of the people. After 8 years of hard work and overcoming a number of difficulties, Ding Liping, a middle-aged scientist at the Iron and Steel Research Institute, and his colleagues finally developed a new type of drill solder with Chinese characteristics used in nuclear engineering. They were among the first group awarded patent rights by the State Patent Bureau and have received large orders totaling a million yuan, creating considerable economic benefit for the state. Hu Guohua, a middle-aged scientist, is doing optical research work for the Ministry of Astronautics. He and another scientist invented a variable optical filter artificial color indicator device, for which a patent was granted. It is a holographic and optical information processing and experiment system, and its invention has provided a new type of research tool for those engaged in optical studies.

Under the warmhearted guidance and help of older experts, Zhao Lei, a young scientist at the Beijing Industrial University, has grown up quickly. He won a first-class prize at the first Edison cup youth invention contest. A small four-way voltage stabilizer he had invented with a lot of hard work was granted a patent for exterior design. A pocket-size transistor on-line automatic testing instrument invented by him has been filed for patent.

12802/12859
CSO: 4008/2107

NATIONAL DEVELOPMENTS

BRIEFS

SINO-FRENCH TECHNOLOGICAL, NUCLEAR ACCORD--Paris, May 23 (XINHUA)--China and France signed a four-year scientific and technological cooperation accord on agronomic research here today. The accord was signed by Lu Liangshu, president of the Chinese Academy of Agricultural Sciences, and Jacques Ploy, director of the French Institute for Agronomic Research. The accord proposes exchanges in the fields of soil microbiology, vegetable planting, rural economy, and improvement of wheat quality. A coordinating committee comprising representatives from both countries will meet every two years in China and France alternately to coordinate the exchanges. Meanwhile, China and France signed a letter of intent here this evening on the evaluations of nuclear safety in a nuclear power plant in Guangdong Province, China. The letter was signed by Jiang Shengjie, director of the National Bureau of Nuclear Safety of China, and Francois Cogne, director of the Nuclear Protection and Safety Institute of France. [Text] [Beijing XINHUA in English 1432 GMT 24 May 86 OW] /6091

ASSOCIATION SEEKS FOREIGN SCHOLARS--Beijing, June 3 (XINHUA)--The China Association for Science and Technology, a non-governmental organization, is trying hard to attract more foreign scholars to come to China for academic exchanges, PEOPLE'S DAILY reported today. In the past six years the association has organized more than 700 international symposiums in China and received more than 30,000 scholars from abroad, according to the paper. The symposiums have helped Chinese scientists know better about the latest developments in science and technology in other countries, the paper said. Meanwhile, the association has sent 3,000 members abroad to study, give lectures, do research or attend international symposiums. More than 200 Chinese scientists have been elected to leading posts of international organizations. [Text] [Beijing XINHUA in English 0759 GMT 3 Jun 86 OW] /6091

CSO: 4010/59

PHYSICAL SCIENCES

MOVING SHOT POINT METHOD AND DATA PROCESSING OF SEISMIC REFRACTION SURVEYS IN SHALLOW SEAS STUDIED

Qingdao SHANDONG HAIYANG XUEYUAN XUEBAO [JOURNAL OF SHANDONG COLLEGE OF OCEANOLOGY] in Chinese Vol 15, No 3, 15 Sep 85 pp 76-84

[Article by Yang Henai [2799 0735 0035], Department of Marine Geology: "The Study of Moving Shot Point Method and Data Processing of Seismic Refraction Surveys in Shallow Sea," paper received on 9 January 1985]

[Text] English Abstract: This paper is a brief summary of engineering geological survey with seismic refraction method over Hongdao, Shijiushuo and Lanshantou shallow water regions during 1976-78. It briefly presents the working method and the interpretation of data which were collected with float cable and continuous profiling in the study area.

Based on operations, the method of data processing and field procedure of seismic refraction method in shallow sea regions were further considered and a theoretical model was set up for data processing.

China's coastal waters are extensive, the sea bottom is rich in natural treasures, and the seacoast has excellent natural harbors. Studying the epigenetic geological structures of the sea bottom has become an important task for developing sea bottom epigenetic ore resources and developing marine engineering. There are mainly three methods of the basic geological surveys which are currently being carried out in the harbors: (1) the exploration drilling method: it can supply the most direct and most reliable geological materials and related technical engineering parameters and it is an indispensable method in marine engineering exploration. Its major shortcomings are: low efficiency, high investment, and it can only provide materials for the well site. (2) Shallow stratigraphic section instrument measurements: This is an advanced technique currently carried out in epigenetic measurement at sea. It is highly efficient, low cost, the materials are directly observed, the strata resolution is high, and it is a method which is widely used today. However, because the signal it emits is high frequency and its power is low, signal energy can decay very quickly in thick strata coarse friable deposits it is very limited in exploration depth. (3) Epigenetic seismic refraction surveys: This is a well-developed and effective exploration method for studying epigenetic geological structures. The energy of its excitation signal is high, frequency is low, exploration depth is deep, and matched with the above two methods it can supply more detailed geological

materials. In 1976-77, the Shandong College of Oceanology's geophysical prospecting adopted this method in harbor engineering geological surveys in planning for the Hongdao reservoir and the Qingdao iron works. Comparing the bedrock depth measured and the materials of many wells, the error was less than 5 percent and excellent geological results were obtained. However, compared to shallow stratigraphic section instrument measurements, its efficiency was low, costs were high and operation at sea was complex and these are the basic reasons why this method has not been widely adopted. In 1978 in selecting harbors on southern Shandong, we successfully tested the single ship continuous operating epigenetic seismic refraction method, raising work efficiency over 20-fold, lowering costs, providing basic depth contour plain view of Shijiuso and Lanshantou harbors which became basic materials for harbor selection and construction in southern Shandong.

This paper summarizes our experience of the past few years in using the epigenetic seismic refraction wave method for harbor engineering geological exploration, discusses the basic principles and observational system of the gas gun and analog or digital tape seismograph for carrying out the continuous operation refraction wave method at sea. In addition, it further presents the interpretation theory and digital processing methods use of moving shot point time interval curve made up of recordings of common [gong 0364] shot points and the theory concerning its procedures and design.

I. The Epigenetic Seismic Refraction Survey Method

A. Geological and Physical Conditions for Epigenetic Seismic Refraction Survey

The physical conditions of seismic refraction method are that the wave velocity in the refraction layer must be greater than the wave velocity in the overlying strata. On the basis of the geological profiles in the harbors we have surveyed, they can be divided basically into upper and lower parts, according to structural and physical analysis. The upper part is weathered crust of igneous rock and water-bearing open deposits of long standing made up of interbedded drift mud, mud, coarse sand, fine sand and silt of unequal thicknesses and in which the sonar wave velocity is between 1500 m/s and 1800 m/s. The lower part is firm igneous rock which has not yet been weathered and in which the sonar wave velocity is between 3000 m/s and 5000 m/s. It is clear that the unweathered igneous rock surface is a clear velocity boundary. Generally a basement stratum of dense structure and higher wave velocity can be found below the open water-bearing deposits. Thus, using the seismic refraction method for carrying out engineering geological surveys in harbors has the necessary geological and physical conditions.

B. Method of Operation at Sea

For the past few years, dynamite explosions have been used at sea as the seismic source with pressure sensitive and piezoelectric detectors receiving the shock information and a domestically manufactured model DZ-663 24 channel magnetic tape seismograph used for recording and reproducing monitor records. The distance between the channels is 10 m, the receiving earth section length is 230 m, and the overall length of the electric cable is 300 m. Two observation systems have been used.

1. The encounter time-interval curve observation system

This system requires that the cable be placed on the sea bottom on the line to be measured and the boats stopped. It requires four 12-30 hp boats to be used as instrument boat (command boat), orientation boat, and two blast boats, respectively. The efficiency of this observation system is low; it can only measure 1.5 km of profile per day, and it is subject to influences of water depth, sea currents, tides, and waves, and such problems that the cable which is placed on the sea bottom cannot be maintained in position along a straight line and thus the shot point positions are not correct.

2. The pursuit time-interval curve observation system

This observation system only requires one 60-hp boat. The instrument group, orientation group, and blast group are all on-board this boat, the detector and the cable are towed from the boat's stern and float at a depth of 7-9 m below the surface. The work is carried out continuously while the boat is traveling at a uniform speed (Fig. 1). While the boat travels along the measurement line at an even speed, an explosion is detonated at fixed time intervals (the time intervals are determined by the distance between shots, the boat's speed, and sea conditions) and at the same time a fix is made on the position of the seismic focus. To maintain a reliable tracking refraction profile, the distance between two shot points should be less than the arrangement length, generally about 150 m. If an airgun seismic focus is used, the distance between shots can be reduced to 30 m, which not only improves the work efficiency, but also more tracks can be realized in the same refraction surface section, thus improving the reliability of the material interpretation. In this way, nearly 1,000 records can be obtained in a day, requiring the use of a computer for digital processing and thus automation of the materials interpretation.

B. Method of materials interpretation

The interpretation of materials is mainly to determine, on the basis of the recorded materials, the normal depth of the refraction surface, the wave velocity in the refraction layer and, using some characteristics in the materials to determine the lithic changes and structural fault situation in the refraction layer.

1. Computing the wave velocity in the refraction layer

When using the encounter time-interval curve observation system, using the normal [zheng 2973] time-interval curve $t_1(x)$ and the reverse [fan 0646] time-interval curve $t_2(x)$, the differential interval curves $t(x)$ and the $t_0(x)$ curve can be determined by E. (1).

$$\left.\begin{aligned} t(x) &= t_1(x) + \Delta t(x) \\ t_0(x) &= t_1(x) - \Delta t(x) \end{aligned}\right\} \qquad (1)$$

In the above equation, $\Delta t(x)=T-t_2(x)$, T represents the interchange time values at shot points 01 and 02 (Fig. 2).

Assuming the wave velocity V_1 in the refraction layer does not change, the pitch of the refraction layer is less than 15°, by conventional computation methods the wave velocity in the refraction layer can be computed using the approximation equation (2).

$$V_1 = 2\frac{dx}{dt} \qquad (2)$$

i.e., the wave velocity in the refraction layer is equal to three times the differential time-interval curve gradient reciprocal.

From Eq. (1) and Fig. 2 it can be seen that the normal time-interval curve of $t_1(x)$ is the time equation bisector of t(x) and $t_o(x)$ and the time, the time at the shot points 01 and 02 are T_{o1} and $2T-t_{o2}$, respectively, the shot interval is L, thus the approximation equation for the wave velocity in the refraction layer can be written

$$V_1 = \frac{2L}{2T-(t_{01}+t_{02})} \qquad (3)$$

When the pursuit time-interval observation system is in continuous operation, Eq. (3) can be used to compute the wave velocity V_1 in the refraction layer. L, T, t_{o1} and t_{o2} in the equation can be solved on the basis of the observation system and recordings.

2. Computing the normal line depth in the refraction layer

When the V_o and V_1 in the received earth section and the boundary gradient does not change, the equation for computing the normal line depth is used

$$h(x) = \frac{V_0}{2\sqrt{1-\left(\frac{V_0}{V_1}\right)^2}}t_0(x) \qquad (4)$$

to find the normal line depth at point x. Here, V_o is the wave velocity in the water and water bearing open depositional layer covering the basement, generally regarded as V_o=1500m/s, V_1 can be used to solve Eqs. (2) and (3), $t_o(x)$ can be found after drawing the time-interval curve.

3. Tide correction

At sea, seismic survey, seismic focus, and detectors are all submerged to a certain depth, thus the boundary normal line depth found at the seismic focus and reception point are influenced by the tides, and in epigenetic geological surveys, when using floating cable, it is necessary to correct for the tide. Since tidal action and the submerged depth of the seismic point and the

detector are different, the seismic focuses and detectors on a measurement line may not be at the same level. Since the submerged depth of the seismic focus and the detectors are different, the depth computed using Eq. (4) is normal line depth from the intermediate horizon surface between the seismic point and the detector to the refraction layer. Tidal action causes the height of the intermediate horizon surface to change with time. To eliminate tidal influence it is necessary to make a tidal correction of the computed depth. Thus, in harbor geological surveys, the tidal reference of zero depth is used to compute the depth H (Fig. 3).

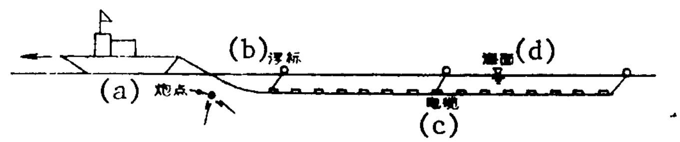


Figure 1. Diagram of Continuous Operation

Key:

a. Shot point
b. Floating marker
c. Electric cable
d. Sea surface

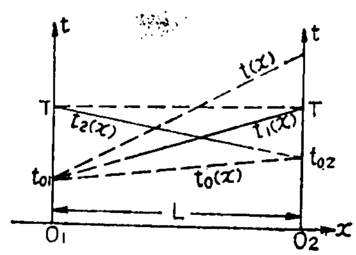


Figure 2. $t(x)$ and $t_o(x)$ Plotted According to $t_1(x)$ and $t_2(x)$

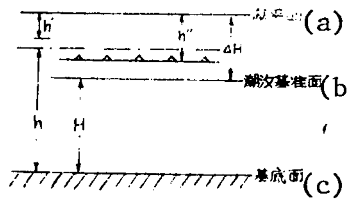


Figure 3. Diagram of Tide Correction

Key:

a. Sea surface
b. Basic tide surface
c. Basement surface

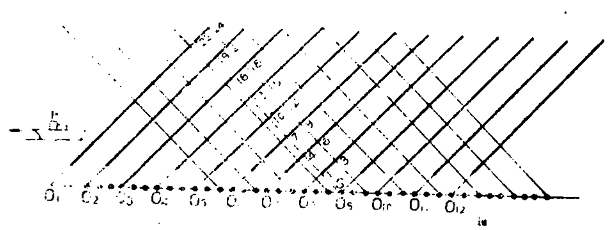


Figure 4. Continuous Pursuit Observation System

The tidal correction equation is

$$H = h - \Delta H + \frac{1}{2}(h' + h'') \qquad (5)$$

in which ΔH is high tide, h' and h" are the submerged depth of the seismic focus and the detectors, respectively.

II. Research on Principles of Moving Shot Point Method Interpretation and Data Processing

In using epigenetic seismic refraction method, carrying out epigenetic geological surveys accurately and at high speed is the key to further development of the refraction method. Based on the experience of operations at sea in the past few years, we recognize that using the single boat continuous operation method and using the moving shot point principles for automated data processing of the materials can greatly improve work efficiency and prospecting accuracy and can open new vistas for epigenetic refraction prospecting.

A. Observation system and method of creating the moving shot point time-interval curve

Since the velocity of the low speed layer at sea is known (1500 m/s), to improve efficiency in operations at sea, the detectors can avoid direct waves by being placed directly in the first arrival zone of the refraction waves. This demands that the horizontal distance x_0 from the first detector to the seismic focus satisfy the equation

$$x_0 \geqslant 2h\sqrt{\frac{V_1 + V_0}{V_1 - V_0}} \tag{6}$$

in which h is the maximum depth of the refraction layer, V_1 is the wave velocity in the refraction stratum, and V_0 is the wave velocity in the low speed layer 1500 m/s.

If we use 24 seismographs and an airgun as a seismic focus, with an interval of 10 m, the boat moving at an even speed of 1 m/s, and an excitation every 30 seconds, we will have continuous tracking observation. From the observation system (Fig. 4) we know that the first seismograph's eighth shot, the fourth seismograph's seventh shot, the seventh seismograph's sixth shot..., the 20-second seismograph's first shot, are at the same reception point S position (i.e., at the ninth shot point). On the basis of the positional relationship between the shot reception points as illustrated in Figure 4, we can find in the eight neighboring common shot point records the eight seismograph records at the same reception point (see Table 1). Using the principle of interchange, the refraction wave time information of these eight seismograph records can be plotted as a refraction wave moving shot point time-interval curve t_2 (Fig. 5). It forms an encounter time-interval curve with the time interval curve $t_1(x)$ of the first shot's common shot point record. Using Eq. (1) we find $t_0(x)$ and on using Eq. (4) compute the refraction layer normal line depth at the reception points.

B. Digital processing method and program design scheme

Digital processing of epigenetic refraction seismic surveys carries out discrete sampling of records obtained in the field, then inputs the sampling values into a computer and carries out the relevant arithmetic processing to determine the initial arrival time when the refraction waves reached each detector, and using the method of least squares finds such parameters as the intercept distance and slope of the refraction wave time-interval curve, computes the wave velocity in the refraction layer and the t_0 values at the reception points, then finally determines the normal line depth of the refraction layer at the reception points. Below we explain the computational method and program outline of the main program.

(b)道号 / 点位号(a) / (c)炮点号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	...
1	22	23	24																										
2	19	20	21	22	23	24																							
3	16	17	18	19	20	21	22	23	24																				
4	13	14	15	16	17	18	19	20	21	22	23	24																	
5	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24														
6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24											
7	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24								
8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24					
9				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
10							1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	...
11										1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	...
12													1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	...
13																1	2	3	4	5	6	7	8	9	10	11	12	13	...
14																			1	2	3	4	5	6	7	8	9	10	...
15																						1	2	3	4	5	6	7	...
...																													
...																													

Table 1. Channel Number Distribution Relationship of Neighboring 8 Shot Records at the Same Point Position

Key:

- a. Point position number
- b. Channel number
- c. Shot point number

1. Extracting the initial arrival time of the refraction wave

This processing is primarily accurately determining the initial arrival time of the refraction waves of each record on the profile to prepare the data for further digital processing. The processing program primarily includes: sending the initial value (such as the number of records on the profile and the initial value of the relevant control variables), inputting a record, preprocessing, finding the initial arrival time of each refraction wave, and storing it in a two-dimensional array G(p,j), up until the records on a profile have all been input, then it enters the follow-up program (see Fig. 6).

Using the cross correlation function, the initial arrival time is computed, accurately measuring the time when the refraction wave arrived at each channel is the key problem to this method, digitally using the cross correlation function to describe quantitatively the relevant degree of two wave forms. It is measuring the degree of similarity of the two waveforms as well as a function of the time displacement between the two waveforms. The cross correlation function is at the maximum when there are only two waveforms overlapped in the process of time displacement change between two identical waveforms. Thus, by computing the cross correlation function, the time difference of the arrival of the refraction wave information at any two detectors can be found.

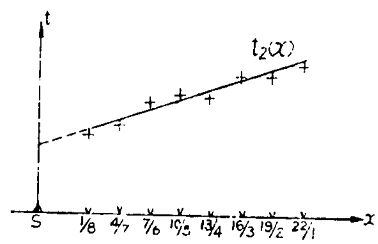
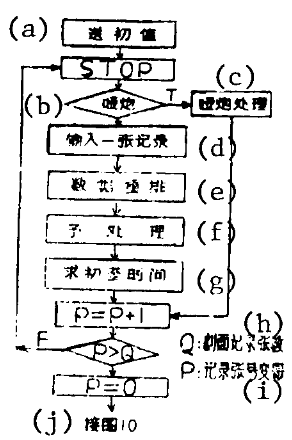


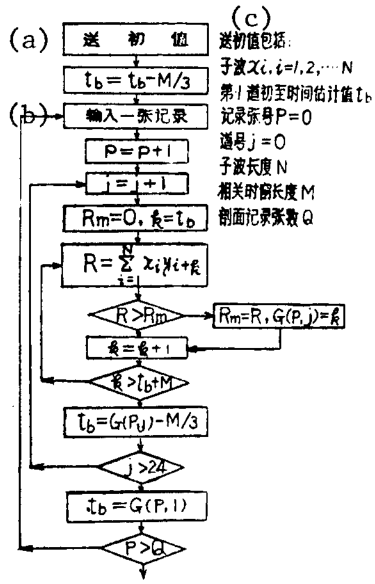
Figure 5. Moving Shot Point Time-Interval Curve



Key:

a. Send initial value
b. Dummy shot
c. Dummy shot processing
d. Input one record
e. Number [unreadable] rearrangement
f. Sub-processing
g. Find initial arrival time
h. Number of profile records
i. Record number variable
j. To Figure 10

Figure 6. Profile Pre-Processing Block Diagram



Key:

a. Send initial value
b. Input one record
c. Send initial value includes:
Wavelet Xi, i=1,2,,...N
First channel initial arrival time t_b
Record number P=0
Channel number j=0
Wavelet length N
Correlation time window length M
Profile record number Q

Figure 7. Block Diagram for Using Cross Correlation to Find Initial Arrival Time on the Channels

If the refraction wave information received from the same source at two receptor channels are x(t) and y(t) respectively, then after discretization, the two groups of time sequences obtained are (x_i), (y_i), i=0,1,2,...,N, their correlation function is

$$R_{xy}(k) = \sum_{i=0}^{N} x_i y_{i+k} \qquad (7)$$

when it constantly changes the k value, we can obtain the maximum value $R_{xy}(kmax)$ of $R_{xy}(k)$, and kmax is the time difference of x_i, y_i. All we have to do is select the refraction wavelet and provide an initial arrival time, and by carrying out cross correlation computations of the information on the various detectors and within the time-wall of a certain length, and we can determine the initial arrival time when the refraction waves reached each channel. The program flowchart for computing the initial arrival time on the channels is illustrated in Figure 7.

2. Equation for finding the time-interval curve

Assuming the boundary inclination and the wave velocity do not change, the time t_i when the refraction wave arrives at each channel and the shot detection distance x_i are correlated linearly. In the rectangular coordinates with time as the longitudinal axis, and distance as the vertical axis, the

dot group (x_i, t_i), i=1,2,...24, reveals a linear distribution, its linear equation is t=a+bx, using the principles of least squares to select the coefficients a and b, its formula is

$$a = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} t_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i t_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \quad (8)$$

$$b = \frac{n\sum_{i=1}^{n} x_i t_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} t_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \quad (9)$$

a is the intercept distance of this time-interval curve, and corresponds to the time t_0 at the shot point. Using known data (x_i, t_i), t=1,2,...24, the program flowchart for finding the time-interval curve coefficients a and b is indicated in Figure 8.

3. Extracting moving shot point time-interval information

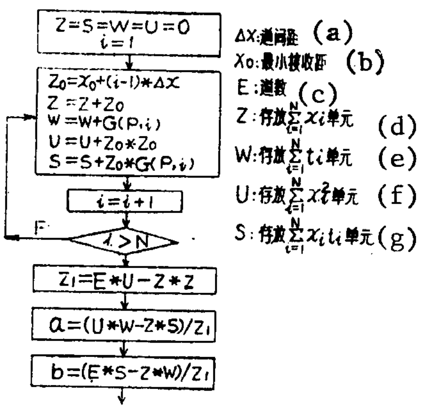
From Fig. 4 and Table 1 one can see that at the reception point after the ninth shot the refraction wave information of eight shots can be obtained, and a moving shot point time-interval curve can be made up and in addition, each additional shot can be obtained from the three moving shot point time-interval curves made up by the previous eight shot points. Under the conditions assumed previously, the distance x from each shot point to the reception point still exhibits a linear relationship with the refraction wave initial arrival time t, and the coefficients a and b of the moving shot point time-interval curve equation are determined according to Eqs. (8) and (9). The key is in extracting from the continuous 8 shot records the initial arrival time on the corresponding channel. Depending on the spatial distribution of the shot points and reception channels in continuous operation, the initial arrival times G(p,j) are extracted using the common shot point records, then the initial arrival time information of the moving shot point records is extracted according to Block Diagram 9.

4. Calculating the wave velocity in the refraction layer and the normal line depth at the reception point

From Fig. 4 it can be seen that starting with the eighth shot, the first-third channel position of each shot record can form three moving shot time-interval curves from the time information of the previous 8 shots. With the time-interval curves of the previous seventh shot, they form three groups of encounter time-interval curves, and the distance of the two coordinates' point of origin is

$$L_n = 30 + (20 + n)\Delta x, \quad n = 1,2,3 \qquad (10)$$

In the equation, Δx is the channel interval (10 m). If after the eighth shot record is input, n=1, $L_{n=1}$ expresses the distance between the two coordinates' point of origin of the first encounter time-interval curve made up of the time interval curve $t_2(x)$ of the first shots common shot point records and the moving shot point time-interval curve $t_1(x)$ at the first channel position of the eighth shot (corresponding to the first shot's 22nd channel position), when n=2, $L_{n=2}$ expresses the distance between the two coordinates' point of origin of the second encounter time-interval curve made up of the first shot's time-interval curve and the moving shot point time-interval curve at the eighth shot's second channel position (corresponding to the first shot's 23rd channel position)...



Key:

a. ΔX: channel intervals
b. X_0: minimum reception distance
c. E: number of channels
d. Z: store cell
e. W: store cell
f. U: store cell
g. S: store cell

Figure 8. Block Diagram for Finding Coefficients a and b

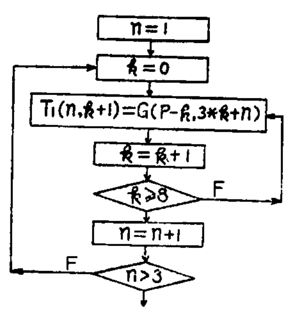


Figure 9. Block Diagram for Extracting Moving Shot Point Initial Arrival Time

Using the method of least squares we solve the coefficients of the two encounter time-interval curve equations. The time-interval curve equation is

$$t = a_{p-7} + b_{p-7} \cdot x \qquad (11)$$

and the fan time-interval curve equation (i.e., the moving shot point time-interval curve equation) is

$$t_n = a'_n + b'_n x \qquad (12)$$

in which a_{p-7} and b_{p-7} represent the time the Pth record is input and the seventh previous shot time-interval curve equation coefficient is solved after the P-7th record is input. a'_n and b'_n represent the $n(1 \leq n \leq 3)$ moving shot point time-interval curve equation coefficient after the input of the Pth record, and can be calculated according to the block diagram in Figure 8, at this time E=8 and Δx=30 m.

(1) Finding wave velocity in refraction layer

Substituting $x=L_n$ in Eqs. (11) and (12) we find the average value of the time as the interchange time value T_n of the encounter time-interval curve, i.e.

$$T_n = \frac{(a'_n + a_{p-7} + b'_n + b_{p-7}) \cdot L_n}{2} \qquad (13)$$

After solving for T_n, then using Eq. (3) we calculate the wave velocity $V_1(n)$ in the refraction layer, i.e.,

$$V_1(n) = \frac{2L_n}{2T_n - (a_{p-7} + a'_n)} \qquad (14)$$

Take the average of the three velocities $V_1(n)$, n=1,2,3, found from the three encounter time-interval curves as the wave velocity in the refraction layer within the p-7 shot reception section, i.e.,

$$\bar{V}_1(p-7) = \frac{1}{3} \sum_{n=1}^{3} V_1(n) \qquad (15)$$

(2) Find the normal line depth of the refraction layer at each reception point on the profile

Using Eq. (1) find the t_0 value of each reception point on the refraction layer normal line depth at each reception point found using the encounter time-interval curve. Suppose x_i is the shot detector distance of the jth channel on the normal time-interval curve, i.e., x_i=30+(j-1)Δx, j=1, 2, ..., 24. In the equation, the number 30 is the minimum shot detector distance, i.e., x_i=30 m, Δx is the interval between channels (10 m).

After inputting the Pth record, calculate the formula for t_o on the j channel position of the P-7th shot as

$$t_0 = a_{p-7} + b_{p-7} \cdot x_j - T_n + [a'_n + b'_n(L_n - x_j)] \tag{16}$$

then substitute t_o into Eq. (4) and calculate the boundary normal line depth at the j channel position.

In addition, it can also be seen in Figure 4 that the reception points after the ninth shot position have recorded the information of the previous 8 shots, and each shot also forms three encounter time-interval curves with the moving shot time-interval curves at channels 21-24, thus beginning with the 8th shot, on the previous three channels of each shot the depth is calculated 24, 23, and 22 times. On channel 21 before the 9th shot position, the depth of 1-3 channel position is calculated 3 times, on 4-6 channel positions, it is calculated 6 times...., and on the 19-21st channel positions it is calculated 21 times. Repeated calculations in this fashion improve the precision of depth calculations. In program design, a sequential number group $H_x(n)$ for each depth calculation point on the profile was created for storing the accumulated value of the depth values that have been repeatedly calculated then the average value of the depth is found depending on the number of accumulations as the final calculated depth. See Figures 10 and 11 for the block diagrams of the calculations.

Figure 11 is the final program block diagram for processing the profile materials. In the figure the dashed line diagram I, for the output profile processing result, to rearrange the accumulated depth values H(p,j) of multiple calculations stored in the reception point positions using the shot point number and channel number in an earlier program into a uniform sequence n according to the reception points on the profile, and put them into the number group $H_x(n)$, $1 \leq n \leq QX3$.

The dashed line box II samples the average depth values of the accumulated depth values $H_x(n)$, $1 \leq n \leq 21$ of the 21 reception points between the second shot position and the ninth shot position according to the number of accumulations at the reception points and sends the calculation result back to the original position $H_x(n)$.

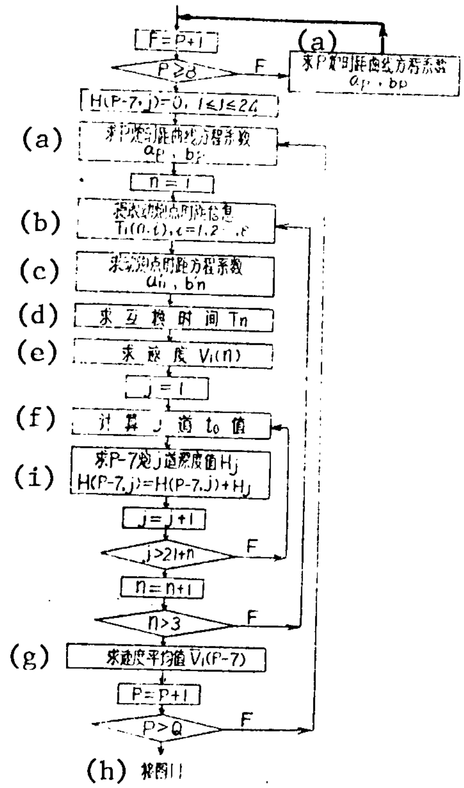


Figure 10. Block Diagram for Finding Velocity and Depth

Key:

a. Find P shot time-interval equation coefficients ap, bp
b. Extract moving shot point time-interval information
c. Find moving shot point time-interval information
d. Find interchange time Tn
e. Find velocity $V_1(n)$
f. Calculate t_o value for channel j
g. Find average value of velocity $V_1(P-7)$
h. To Figure 11
i. Find P-7 shot j channel depth Hj

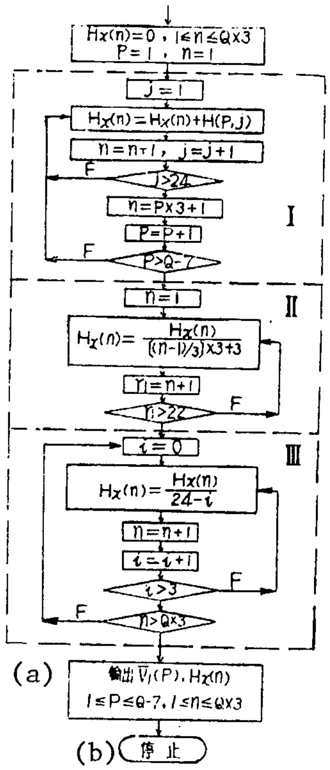


Figure 11. Output Velocity and Depth Block Diagram

Key:

a. Output
b. Stop

Dashed line III calculates the depth of the reception points ($n>21$) after the ninth shot position on the profile. Because each three neighboring point positions have accumulations of 24, 23, and 22 depth calculation values, it is necessary to average the depth values of the number of actual accumulations, and then send it back to the original position $H_x(n)$, $21<n\leq Qx3$.

After the depth values at all the reception points on the profile have been extracted, they are output according to the reception point sequence number n, and at the same time the velocity $\overline{V}_1(P)$, $1\leq P\leq Q-7$ is output according to the shot point sequence number P. Depending on changes in the output values, the boundary depth and lateral changes in velocity can be understood and the structural state of the refraction layer along the profile can be further deduced, this concludes the profile processing. Similar processing can be carried out for other profiles, but due to differences in the measurement line length, the initial conditions at processing time are not the same.

8226/12712
CSO: 4007/1054

PHYSICAL SCIENCES

APPLICATION OF FAST HILBERT TRANSFORM CALCULATIONS TO INTERPRETATION OF TWO-DIMENSIONAL MAGNETIC ANOMALIES

Qingdao SHANDONG HAIYANG XUEYUAN XUEBAO [JOURNAL OF SHANDONG COLLEGE OF OCEANOLOGY] in Chinese Vol 15, No 3, 15 Sep 85 pp 67-75

[Article by Wang Shuoru [3762 4311 0320] and Yu Tao [0060 3447], Department of Marine Geology: "Fast Hilbert Transform Calculations and Their Application to the Interpretation of Two-Dimensional Magnetic Anomalies"; paper received 6 November 1984]

[Text] Abstract: Hilbert transform is a method for digital processing. This paper presents two types of Hilbert transform formulas and their calculation formulas in frequency domain based on Fourier transforms. Hilbert transform formulas of $\Delta\tilde{T}_1$ and $\Delta\tilde{T}_2$ of the total magnetic effect of some two-dimensional magnetic bodies are obtained according to the correlation of component magnetic effect. A new procedure for ΔT interpretation is developed using some special points and the intersection of $\Delta\tilde{T}_1$, $\Delta\tilde{T}_2$ and ΔT curves. The validity of the procedure is tested on theoretical models and a field example.

The basic task of gravity and magnetic [zhongci 6850 4318] data processing is to extract more information from observation data and by a method of enriching the interpretation of gravity and magnetic anomalies, to improve the accuracy and reliability of the results of geological interpretation. Hilbert transforms are a method of data processing which in recent years have been used in gravity and magnetic anomalies. Hilbert transforms have been used primarily in conversion [husuan 0062 4615][1]; and to reflect through some components and the Hilbert transform of a two-dimensional body's gravity and magnetic anomaly the essentials of hidden depth and occurrence[2,3]. Undoubtedly, these results have a definite practical significance. However, whether or not Hilbert transforms have a relationship only with place field component conversion needs to be further clarified. Next, what is obtained from magnetometry on the ocean surface is the total magnetic [zongci 4920 4318] anomaly ΔT, thus, what is even more practical is to use Hilbert transforms directly on ΔT to interpret the anomaly. From the practical angle, it is also very important to determine a simpler method of Hilbert transform. These three items are the main problems discussed in this paper.

I. Expression of Hilbert Transform

The definition of the Hilbert transform $\tilde{g}(x)$ of a function $g(x)$ is[4]

$$\tilde{g}(x)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{g(\tau)}{x-\tau}d\tau \text{ 。} \tag{1}$$

Its contratransform is

$$g(x)=-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\tilde{g}(\tau)}{x-\tau}d\tau \text{ 。} \tag{2}$$

The faltung form of Eq. (1) is written

$$\tilde{g}(x)=\frac{1}{\pi x}*g(x) \text{ 。}$$

Correspondingly

$$\tilde{G}(\omega)=H(\omega)\cdot G(\omega) \tag{3}$$

in which $\tilde{G}(w)$, $G(w)$ are the Fourier transforms of $\tilde{g}(x)$ and $g(x)$ respectively, and $H(w)$ is the frequency spectrum of $\frac{1}{\pi x}$.

If the Fourier transform is defined as

$$F(\omega)=\int_{-\infty}^{\infty}f(x)\cdot e^{-j\omega x}dx \text{ ,}$$

because

$$H(\omega)=\begin{cases}-j & \omega>0 \\ j & \omega<0 \text{ ,}\end{cases}$$

then Eq. (3) becomes

$$\tilde{G}(\omega) = \begin{cases} -j \cdot G(\omega) & \omega > 0 \\ j \cdot G(\omega) & \omega < 0 \end{cases} \qquad (4)$$

Let $$G_1(\omega) = G_R(\omega) + j \cdot G_I(\omega) \qquad (5)$$

then from Eq. (4) we can obtain

$$\tilde{G}_1(\omega) = \begin{cases} G_I(\omega) - j \cdot G_R(\omega) & \omega > 0 \\ -G_I(\omega) + j \cdot G_R(\omega) & \omega < 0 \end{cases} \qquad (6)$$

For the real function, the real part of its Fourier transform is the accidental function, and the imaginary part is the odd function. Therefore we should have

$$G_1(-\omega) = G_R(\omega) - j \cdot G_I(\omega) \qquad (7)$$

The contra-Fourier transform of $\tilde{G}_1(w)$ is

$$\begin{aligned} \tilde{g}_1(x) &= \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} H(\omega) \cdot G_1(\omega) \cdot e^{j\omega x} d\omega \\ &= \frac{1}{2\pi} \left[\int_{-\infty}^{0} j \cdot G_1(\omega) \cdot e^{j\omega x} d\omega - \int_{0}^{\infty} j \cdot G_1(\omega) \cdot e^{j\omega x} d\omega \right] \\ &= \frac{1}{2\pi} \left[\int_{0}^{\infty} j \cdot G_1(-\omega) \cdot e^{-j\omega x} d\omega - \int_{0}^{\infty} j \cdot G_1(\omega) e^{j\omega x} d\omega \right] \end{aligned}$$

Substitution the two equations (5) and (7) and collating them we get

$$\tilde{g}_1(x) = \frac{1}{\pi} \int_{0}^{\infty} [G_R(\omega) \cdot \sin\omega x + G_I(\omega) \cdot \cos\omega x] d\omega \qquad (8)$$

If we let

$$G_2(\omega) = -G_R(\omega) + jG_I(\omega) , \qquad (9)$$

then based on Eq. (4), we should have

$$\tilde{G}_2(\omega)=\begin{cases}G_I(\omega)+j\cdot G_R(\omega) & \omega>0\\ -G_I(\omega)-j\cdot G_R(\omega) & \omega<0\,.\end{cases} \tag{10}$$

The derivational process for Eq. (8) is similar and we can obtain the contra-transform of $\tilde{G}$ (w)

$$\tilde{g}_2(x)=\frac{1}{\pi}\int_0^{\infty}[-G_R(\omega)\cdot\sin\omega x+G_I(\omega)\cdot\cos\omega x]d\omega\,. \tag{11}$$

From this it can be seen that we can provide two different definitions of the Hilbert transform, i.e., Eq. (8) and Eq. (11) (to distinguish the two we will use different subscripts for the two transforms).

It has already been proven that for two dimensional space, the two equations (6) and (8) are the interrelationship of the orthogonal components in a single plane place field. The Hilbert transform $\Delta\tilde{H}_1(x)$ of the surface magnetic anomaly $\Delta H(x)$ is $\Delta Z(x)$, the Hilbert transform $\Delta\tilde{Z}_1(z)$ of the magnetic measurement $\Delta Z(z)$ in a straight hole corresponds to $\Delta H(z)$; but the contra-Hilbert transform of $\Delta Z(x)$ and $\Delta H(z)$ corresponds to $H(x)$ and $Z(z)$ [1].

II. The Hilbert Transforms $\Delta T_1(x)$ and $\Delta T_2(x)$ of the $\Delta T(x)$ of Some Regular Two-Dimensional Bodies

According to the definition of Hilbert transforms given above, we can find the Hilbert transform formula of two-dimensional total magnetic anomalies $\Delta T(x)$.

A. Method of finding Hilbert transform of $\Delta T(x)$

Since the $\Delta T(x)$ of horizontal columnar body corresponding to i_o changes to $\Delta H(x)$; the $\Delta T(x)$ of a plate-shaped body corresponding to γ angle changes to the $\Delta H(x)$ when $(\gamma-i_o)$ is the angle. The i_o here is the effective magnetization inclination; γ is the included angle of the magnetization direction and the plate surface; $\gamma=\alpha-i_o$ (α is the inclination of the plate). Therefore, according to $\Delta\tilde{H}_1(x)=\Delta Z(x)$, it is very easy to obtain $\Delta\tilde{T}_1(x)$.

According to the odd accidence of Fourier transform, taking the even item sign reversal [oucixiang bianhao 0260 3945 7309 6239 5714] in $\Delta\tilde{T}_1(x)$ one can immediately obtain $\Delta\tilde{T}_2(x)$.

B. Hilbert transform expression of some regular two-dimensional bodies

According to the above method, it is not difficult to write the Hilbert transform equations of regular megnetic two-dimensional bodies.

1. Horizontal columnar body

$$\Delta T(x)=\frac{A}{(x^2+h^2)^2}[-(h^2-x^2)\cos(2i_0)-2hx\sin(2i_0)]\ 。\tag{12}$$

$$\Delta \tilde{T}_1(x)=\frac{A}{(x^2+h^2)^2}[(h^2-x^2)\cdot\sin(2i_0)-2hx\cos(2i_0)]\ 。\tag{13}$$

$$\Delta \tilde{T}_2(x)=\frac{A}{(x^2+h^2)^2}\cdot[-(h^2-x^2)\cdot\sin(2i_0)-2hx\cos(2i_0)]\ 。\tag{14}$$

2. Thin plate of infinite extension

$$\Delta T(x)=\frac{B}{x^2+h^2}\cdot[-h\cdot\sin(\gamma-i_0)-x\cdot\cos(\gamma-i_0)]\ 。\tag{15}$$

$$\Delta \tilde{T}_1(x)=\frac{B}{x^2+h^2}\cdot[h\cos(\gamma-i_0)-x\cdot\sin(\gamma-i_0)]\ 。\tag{16}$$

$$\Delta \tilde{T}_2(x)=\frac{B}{x^2+h^2}[-h\cos(\gamma-i_0)-x\cdot\sin(\gamma-i_0)]\ 。\tag{17}$$

3. Thick plate of infinite extension

$$\Delta T(x)=C\cdot\left[-\mathrm{tg}^{-1}\frac{2bh}{x^2-b^2+h^2}\cdot\sin(\gamma-i_0)+\frac{1}{2}\ln\frac{(x+b)^2+h^2}{(x-b)^2+h^2}\cdot\cos(\gamma-i_0)\right]\ 。\tag{18}$$

$$\Delta \tilde{T}_1(x)=C\cdot\left[\mathrm{tg}^{-1}\frac{2bh}{x^2-b^2+h^2}\cdot\cos(\gamma-i_0)+\frac{1}{2}\ln\frac{(x+b)^2+h^2}{(x-b)^2+h^2}\cdot\sin(\gamma-i_0)\right]\ 。\tag{19}$$

$$\Delta \tilde{T}_2(x)=C\cdot\left[-\mathrm{tg}^{-1}\frac{2bh}{x^2-b^2+h^2}\cos(\gamma-i_0)+\frac{1}{2}\ln\frac{(x+b)^2+h^2}{(x-b)^2+h^2}\sin(\gamma-i_0)\right]\ 。\tag{20}$$

4. Horizontal thin plate

$$\Delta T(x)=A\cdot\frac{[-(h^2+b^2-x^2)\cos(2i_0)-2hx\sin(2i_0)]}{[h^2+(x+b)^2]\cdot[h^2+(x-b)^2]}\ 。\tag{21}$$

$$\Delta \tilde{T}_1(x)=A\cdot\frac{[(h^2+b^2-x^2)\sin(2i_0)-2hx\cos(2i_0)]}{[h^2+(x+b)^2]\cdot[h^2+(x-b)^2]}\ 。\tag{22}$$

$$\Delta \tilde{T}_2(x)=A\cdot\frac{[-(h^2+b^2-x^2)^2\cdot\sin(2i_0)-2hx\cos(2i_0)]}{[h^2+(x+b)^2]\cdot[h^2+(x-b)^2]}\ 。\tag{23}$$

5. Thin plate of finite extension

$$\Delta T(x) = B\cdot\left[\frac{-h_1\sin(\gamma-i_0)-x\cos(\gamma-i_0)}{x^2+h_1^2}+\frac{x\cdot\cos(\gamma-i_0)+h_2\cdot\sin(\gamma-i_0)}{x^2+h_2^2}\right]。 \quad (24)$$

$$\Delta\tilde{T}_1(x) = B\cdot\left[\frac{h_1\cos(\gamma-i_0)-x\sin(\gamma-i_0)}{x^2+h_1^2}+\frac{x\sin(\gamma-i_0)-h_2\cos(\gamma-i_0)}{x^2+h_2^2}\right]。 \quad (25)$$

$$\Delta\tilde{T}_2(x) = B\cdot\left[\frac{-h_1\cos(\gamma-i_0)-x\sin(\gamma-i_0)}{x^2+h_1^2}+\frac{x\sin(\gamma-i_0)+h_2\cos(\gamma-i_0)}{x^2+h_2^2}\right]。 \quad (26)$$

In which, A is the constant related to the intensity of magnetization; B is the constant related to the intensity of magnetization, plate width and plate inclination; C is the constant related to intensity of magnetization and plate inclination; b is the half-width of the plate; h_1 and h_2 are the thin plate top surface and bottom surface depth, respectively.

The horizontal columnar body and the plate body are representative two-dimensional bodies, and the two-dimensional body of any cross-section can be viewed as a combination of many plate bodies or many horizontal columnar bodies. Thus, the above relational expression in principle is also suited to any two-dimensional body.

III. A Simple and Fast Algorithm for Implementing the Hilbert Transform

Calculating the Hilbert transforms $\tilde{g}_1(x)$ and $\tilde{g}_2(x)$ of the function $g(x)$ can be carried out using the formulas (8) and (11), and their special functions $\tilde{G}_1(w)$ and $\tilde{G}_2(w)$ can also first be carried out according to formulas (6) and (10) and then the Fourier contratransform can be carried out to find $g_1(x)$ and $g_2(x)$. The second algorithm is simpler. The specific steps are as follows:

1. Use FFT to find the Fourier contratransform $G(x)$ of the original function $g(x)$.

2. Change the sign of the data of the real part $w>0$ and the imaginary part $w<0$ of $G(w)$ and then interchange the real and imaginary parts, to obtain $\tilde{G}_1(w)$; change the signs of the data of the real part and imaginary part of $w<0$ of $G(w)$, then interchange the real part and the imaginary part to obtain $\tilde{G}_2(w)$.

3. Use FFT for the contra Fourier transforms, and after transform, $\tilde{g}_1(x)$ is obtained from $\tilde{G}_1(w)$ and $\tilde{g}_2(x)$ from $\tilde{G}_2(w)$.

IV. Applying Hilbert transforms Direct Inversion $\Delta T(x)$

On the basis of the intersecting coordinates of the three curves of $\Delta T(x)$, $\Delta\tilde{T}_1(x)$ and $\Delta\tilde{T}_2(x)$, the hidden depth and occurrence elements in some regular two-dimensional bodies can be inverted directly.

1. Horizontal columnar body

On the basis of the two Eqs. (13) and (14), letting $\Delta\tilde{T}_1(x) = \Delta\tilde{T}_2(x)$, we can get

$$h = \Delta x_{1,2}/2 \text{。} \tag{27}$$

in which $\Delta x_{1,2}$ is the horizontal space of the intersection of $\Delta\tilde{T}_1(x)$ and $\Delta\tilde{T}_2(x)$.

Using $\Delta\tilde{T}_1(0)/\Delta T(0)$ (or $\Delta\tilde{T}_2(0)/\Delta T(0)$) we can get

$$i_0 = -\frac{1}{2}\mathrm{tg}^{-1}\frac{\Delta\breve{T}_1(0)}{\Delta T(0)}\left(\text{或}\ i_0 = \frac{1}{2}\mathrm{tg}^{-1}\frac{\Delta\breve{T}_2(0)}{\Delta T(0)}\right) \text{。}$$

or

$$i_0 = \frac{1}{4}\left(\mathrm{tg}^{-1}\frac{\Delta\breve{T}_2(0)}{\Delta T(0)} - \mathrm{tg}^{-1}\frac{\Delta\breve{T}_1(0)}{\Delta T(0)}\right) \text{。} \tag{28}$$

and

$$(\Delta T^2(0) + \Delta\breve{T}_1^2(0))^{1/2} = A/h^2 ,$$

therefore

$$A = h^2(\Delta\breve{T}_1^2(0) + \Delta T^2(0))^{1/2} \text{。} \tag{29}$$

2. Thin plate of infinite extension

By the same principle, from $\Delta T(x)=\Delta\tilde{T}_2(x)$, we can get

$$h = x_0 \text{。} \tag{30}$$

in which x_o is the horizontal coordinate of $\Delta T(x)$ and $\Delta\tilde{T}_2(x)$.

From $\Delta T(x)/\Delta\tilde{T}_2(x)$ we can get

$$\gamma - i_0 = tg^{-1}\frac{\Delta T(0)}{\Delta \tilde{T}_2(0)}\left(或 - tg^{-1}\frac{\Delta T(0)}{\Delta \tilde{T}_1(0)}\right)。 \qquad (31)$$

and

$$B = h\cdot(\Delta T^2(0) + \Delta\tilde{T}_1^2(0))^{\frac{1}{2}} 。 \qquad (32)$$

3. Thick plate of infinite extension

For Eq. (18) and Eq. (20) we derive, respectively

$$\Delta T'(x) = C\cdot\Big[\sin(\gamma - i_0)\Big(\frac{-h}{(x+b)^2+h^2} + \frac{h}{(x-b)^2+h^2}\Big) + \cos(\gamma - i_0)\Big(\frac{x+b}{(x+b)^2+h^2} - \frac{x-b}{(x-b)^2+h^2}\Big)\Big],$$

$$\Delta \tilde{T}_2'(x) = C\cdot\Big[\cos(\gamma - i_0)\Big(\frac{-h}{(x+b)^2+h^2} + \frac{h}{(x-b)^2+h^2}\Big) + \sin(\gamma - i_0)\Big(\frac{x+b}{(x+b)^2+h^2} - \frac{x-b}{(x-b)^2+h^2}\Big)\Big]。$$

Letting $\Delta T'(x)=\Delta\tilde{T}_2'(x)$, we can obtain the following equation:

$$x^2+2hx-h^2-b^2=0$$

solution, $x_{1,2}=-h\pm\sqrt{2h^2-b^2}$.

From this we can solve

$$h = -(x_1+x_2)/2 。 \qquad (33)$$

$$b = \left[\frac{(x_1-x_2)^2-2(x_1+x_2)^2}{4}\right]^{1/2} 。 \qquad (34)$$

In which x_1 and x_2 are the horizontal coordinates of the two intersecting points $\Delta T'(x)$ and $\Delta\tilde{T}_2'(x)$. Furthermore, we can easily solve

$$\gamma - i_0 = tg^{-1}\frac{\Delta T(0)}{\Delta \tilde{T}_2(0)} 。 \qquad (35)$$

$$C = -\frac{h^2+b^2}{2b}(\Delta T'^{2}(0) + \widetilde{\Delta T}_2'^{2}(0))^{1/2} \quad \text{。} \tag{36}$$

4. Horizontal thin plate

From $\Delta\tilde{T}_1(x)=\Delta\tilde{T}_2(x)$ we get the solution $x_{1,2}=\pm(h^2+b^2)^{1/2}$, thus the difference of the two roots

$$\Delta x_{12}=2\cdot(h^2+b^2)^{1/2}. \tag{37}$$

from $\Delta T(x)=\Delta\tilde{T}_2(x)$, we get the solution $x_{1,2}=-h\pm(2h^2+b^2)^{1/2}$, thus the difference of the two roots

$$\Delta x_{02}=2\cdot(2h^2+b^2)^{1/2}. \tag{38}$$

From Eqs. (37) and (38) we can solve

$$h = (\Delta x_{02}^{2} - \Delta x_{12}^{2})^{1/2}/2 \ , \tag{39}$$

$$b = (2\Delta x_{12}^{2} - \Delta x_{02}^{2})^{1/2}/2 \quad \text{。} \tag{40}$$

and, easily solve

$$i_0 = \frac{1}{2}\,\mathrm{tg}^{-1}\frac{\widetilde{\Delta T}_2(0)}{\Delta T(0)} \quad \text{。} \tag{41}$$

$$A = (h^2+b^2)\cdot(\Delta T^2(0) + \widetilde{\Delta T}_1^{2}(0))^{1/2} \quad \text{。} \tag{42}$$

5. Thin plate of finite extension

By the same method described above, letting $\Delta T(x)=\Delta\tilde{T}_2(x)$, $\Delta\tilde{T}_1(x)=\Delta\tilde{T}_2(x)$, it is not difficult to solve

$$h_1 = [(\Delta x_{0,2})^2 - \Delta x_{1,2}^{2})^{1/2} - (\Delta x_{0,2}^{2} - 2\Delta x_{1,2}^{2})^{1/2}]/2 \quad \text{。} \tag{43}$$

$$h_2 = [(\Delta x_{0,2}^{2} - \Delta x_{1,2}^{2})^{1/2} + (\Delta x_{0,2}^{2} - 2\Delta x_{1,2}^{2})^{1/2}]/2 \quad \text{。} \tag{44}$$

In which, $\Delta x_{0,2}$ and $\Delta x_{1,2}$ represent the horizontal space of the intersecting points of curves $\Delta T(x)$ and $\Delta\tilde{T}_2(x)$ and $\Delta\tilde{T}_1(x)$ and $\Delta\tilde{T}_2(x)$.

Furthermore, it is easy to solve

$$\gamma - i_0 = \mathrm{tg}^{-1}\frac{\Delta T(0)}{\Delta\tilde{T}_2(0)} \quad (45)$$

$$B = \frac{h_1 h_2}{h_2 - h_1}(\Delta T^2(0) + \Delta\tilde{T}_1^2(0))^{1/2} \quad (46)$$

It must be pointed out that the form of expression of the contraproblem solutions are not uniform. Solutions in other forms can be obtained depending on the other characteristic points of the curves.

As we said above, when applying the magnetic anomaly curve and its Hilbert transform curve for inversion, the starting point of the coordinates must be determined. Since $\Delta\tilde{T}_1(x)=-\Delta\tilde{T}_2(-x)$, it is most convenient to use this asymmetrical relationship to determine the starting point. The specific method is to first find the intersection o1 and o2 of the two curves $\Delta\tilde{T}_1(x)$ and $\Delta\tilde{T}_2(x)$, then the central point on the line segment o1o2 is the coordinate starting point.

Another method for determining the starting point is to find the maximum value of $(\Delta T^2(x)+\Delta\tilde{T}_1^2(x))^{1/2}$, and the maximum value is the coordinate starting point. However, this method is rather troublesome. It should also be pointed out that the maximum value point of $[\Delta T^2(x)+\Delta\tilde{T}_2^2(x)]^{1/2}$ is not the coordinate starting point (this is because the signal made up of $\Delta T+j\Delta\tilde{T}_2$ is not an analytical signal. In addition, this can also be seen directly from the theoretical form of regular bodies). Thus, N.L. Mohan and others used the method of determining the maximum value point by solving $(\Delta Z^2(x)+\Delta\tilde{Z}_2^2(x))^{1/2}$[3] which is theoretically in error.

V. Theoretical Model and Examples

Using Hilbert transform, the authors made experimental calculations on two models: a horizontal columnar body and a horizontal thin plate, and found the inversion result and the carried out inversion interpretation of a field example.

Figures 1 and 2 are respectively the magnetic anomaly theoretical curve and the theoretical and calculation curve of its Hilbert transform on the above model when taking 64 sampling points. Tables 1 and 2 give the horizontal columnar body and horizontal thin plate model parameters and their inversion results.

From Tables 1 and 2 it can be seen that although the sampling number was small, the inversion precision was very high.

Figure 3 is a magnetic measurement profile of a place in Shandong. Actually, the ore body is a lens inclined toward the north, the downward length is smaller than the central submerged depth, it can be seen that it is nearly a horizontal columnar body[5]. On the basis of the above described interpretation method, using Figure 4 we found the inversion results: h=30 m, i_o=39°. The interpretation results of authors of reference [5] who used the tangent method were: h=413 m, i_o=48°. Number 2 test drilling found ore at 336 m. From this it can be seen that precision of this interpretation method is higher than the tangent method.

Table 1. Comparison of Horizontal Columnar Body Parameter Value Inversion Results

Model parameter	Theoretical value	Calculated Value		Error	
		64 dot sample	128 dot sample	64 dot sample	128 dot sample
h	200	186.9	197.5	7.55%	1.25%
i_o	30°	33.3°	30.5°	11.1%	1.8%
$A \times 10^7$	1.0	0.8314	0.9864	16.9%	1.36%

Table 2. Comparison of Horizontal Thin Plate Paramater Value Inversion Results

Model parameter	Theoretical value	Calculated Value		Error	
		64 dot sample	128 dot sample	64 dot sample	128 dot sample
h	110.0	114.2	108.9	3.8%	1.0%
b	59.0	61.0	57.0	3.39%	3.39%
i_o	78°	80.0°	79.0°	2.6%	1.3%
$A \times 10^7$	1.0	1.0726	0.9689	7.3%	3.1%

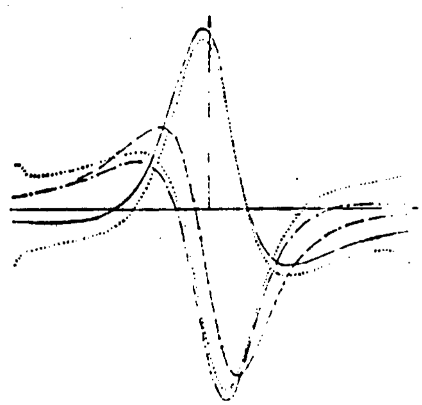


Figure 1. Theoretical and Calculated Curves of Horizontal Columnar Body

[Key on following page]

Key:

a. Dashed line ΔT theoretical curve
b. Solid line $\Delta\tilde{T}_1$ theoretical curve
c. Dot-dash line $\Delta\tilde{T}_2$ theoretical curve
d. Circle line calculated results of $\Delta\tilde{T}_1$ and $\Delta\tilde{T}_2$

Field value is any unit

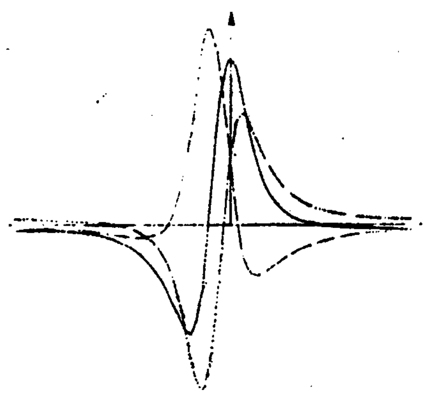


Figure 2. Theoretical and Calculated Curves of Horizontal Thin Plate Body

Key:

a. Dashed line ΔT theoretical curve
b. Solid line $\Delta\tilde{T}_1$ theoretical curve
c. Dot-dash line $\Delta\tilde{T}_2$ theoretical curve
d. Circle line calculated results of $\Delta\tilde{T}_1$ and $\Delta\tilde{T}_2$

Field value is any unit

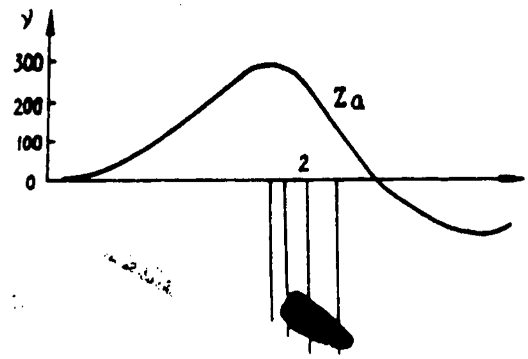


Figure 3. Profile of a Measured Magnetic Anomaly in Shandong

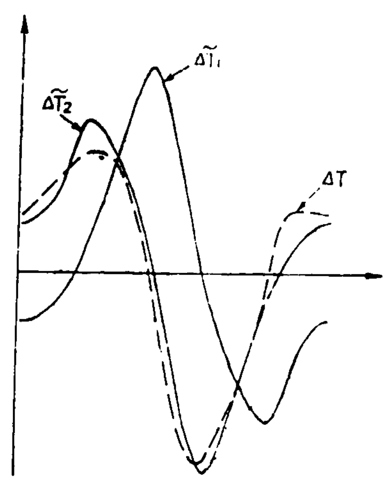


Figure 4. ΔT, $\Delta\tilde{T}_1$, $\Delta\tilde{T}_2$ Curves Calculated According to Measured Curves

Summarizing the above, since the inversion method presented in this paper uses the fast Hilbert transform method, it is simple and easy to use and is suited to applications in field conditions with a microcomputer.

REFERENCES

1. Wang Shuoru [3769 4311 0320], "A Two-dimensional Place Component Husan in the Frequency Domain," Shandong HAIYANG XUEYUAN XUEBAO [JOURNAL OF THE SHANDONG COLLEGE OF OCEANOLOGY], No 4, 1984.

2. R. Green, "Using the Gravitational Method to Determine Precisely the Inclination of a Geological Contact Surface," GUOWAI WUTAN YIWENJI, Kexue Jishu Wenxian Chubanshe, 1978.

3. N.L. Mohan, N. Sundarajian and S.V. Seshagirirao, Interpretation of Some Two-dimensional Magnetic Bodies Using Hilbert Transform Geophysics, No 3, 1982.

4. Cheng Qiansheng [4453 0051 3932], XINHAO SHUZI CHULIDE SHUXUE YUANLI [MATHEMATICAL PRINCIPLES OF DIGITAL SIGNAL PROCESSING], Shiyou Gongye Chubanshe, 1979.

5. Liang Jinwen [2733 6930 2429], SHUIPING YUANZHUTI CI YICHANG DINGLIANG JIESHIDE JIANYI FANGFA [A SIMPLE METHOD OF HORIZONTAL COLUMNAR BODY MAGNETIC ANOMALY QUANTITATIVE INTERPRETATION], Dizhi Yu Kantan, No 11, 1980.

8226/12712
CSO: 4008/1054

PHYSICAL SCIENCES

INVERSION OF CIRCULANT MATRIX DISCUSSED

Chongqing YINGYONG SHUXUE HE LIXUE [APPLIED MATHEMATICS AND MECHANICS] in Chinese Vol 5 No 5, Sep 84 pp 683-690

[Article by Wang Yingjian [3769 4481 1017], Department of Mechanics, Beijing University; paper received 9 August 1983]

[Text] Abstract: A method for finding the inverse of a circulant matrix is described and the representation of the inverse is presented.

I. Introduction

For cubic spline interpolation on N + 1 points, $x_0<x_1<x_2<\cdots<x_N$ with periodic boundary conditions, the following group of linear equations is derived [1]:

$$\begin{bmatrix} 2 & \mu_1 & 0 & \cdot & \cdot & \cdot & 0 & 0 & \lambda_1 \\ \lambda_2 & 2 & \mu_2 & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ 0 & \lambda_3 & 2 & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \ddots & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \ddots & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \ddots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 2 & \mu_{N-2} & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \lambda_{N-1} & 2 & \mu_{N-1} \\ \mu_N & 0 & 0 & \cdot & \cdot & \cdot & 0 & \lambda_N & 2 \end{bmatrix} \begin{Bmatrix} y'_1 \\ y'_2 \\ y'_3 \\ \cdot \\ \cdot \\ \cdot \\ y'_{N-2} \\ y'_{N-1} \\ y'_N \end{Bmatrix} = \begin{Bmatrix} q_1 \\ q_2 \\ q_3 \\ \cdot \\ \cdot \\ \cdot \\ q_{N-2} \\ q_{N-1} \\ q_N \end{Bmatrix} \tag{1.1}$$

Where $y'_1, y'_2, \cdots, y'_N$ are the first derivatives at the interpolation points, and

$$\lambda_k = \frac{h_{k+1}}{h_k+h_{k+1}}, \quad \mu_k = 1-\lambda_k \qquad (k=1,2,\cdots, N-1)$$

$$\lambda_N = \frac{h_1}{h_N+h_1}, \quad \mu_N = 1-\lambda_N, \quad h_k = x_k - x_{k-1} \qquad (k=1,2,\cdots,N)$$

on the right side of equation (1.1),

$$q_k = 3\lambda_k \frac{y_k - y_{k-1}}{h_k} + 3\mu_k \frac{y_{k+1} - y_k}{h_{k+1}} \qquad (k=1,2,\cdots,N)$$

where $y_k (k=1,2,\cdots,N)$, $y_{N+1}=y_1$ are the formula values at the interpolation points. Equation (1.1) can be solved by a method similar to "pursuit method" [1].

But for cubic spline interpolation on equidistant nodes with periodic boundary conditions, we derive the linearized algebraic equations

$$\begin{bmatrix} 4 & 1 & 0 & \cdot & \cdot & \cdot & 0 & 0 & 1 \\ 1 & 4 & 1 & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ 0 & 1 & 4 & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \ddots & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \ddots & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \ddots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 4 & 1 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 & 4 & 1 \\ 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 & 4 \end{bmatrix} \begin{Bmatrix} y'_1 \\ y'_2 \\ y'_3 \\ \cdot \\ \cdot \\ \cdot \\ y'_{N-2} \\ y'_{N-1} \\ y'_N \end{Bmatrix} = \begin{Bmatrix} p_1 \\ p_2 \\ p_3 \\ \cdot \\ \cdot \\ \cdot \\ p_{N-2} \\ p_{N-1} \\ p_N \end{Bmatrix} \tag{1.2}$$

which we write

$$A\{X\}=\{P\}$$

where $\{X\}=(y'_1, y'_2, \cdots, y'_N)^T$, $\{P\}=(p_1, p_2, \cdots, p_N)^T$, and matrix A is a circulant matrix.

Below we discuss the method of inversion of the circulant matrix and give a formula for the inverse. Using the results, we can immediately obtain the inverse A^{-1} of matrix A in equation (1.2), which simplifies the process of solving this equation.

Structural calculations using circulant matrices have important applications; the inversion method given here for them is applicable to such calculations.

The discussion below is an extension of that in Ref. 2.

II. An Important Property of the Circulant Matrix

A partitioned matrix of order m ($m \geq 3$) can be represented as

$$A=(A_{uv}) \qquad (u, v=0,1,\cdots, m-1)$$

where the submatrices A_{uv} are n x n square matrices. Accordingly, A is an (m x n) x (m x n) matrix.

If

$$A=\begin{bmatrix} A_0 & A_1 & A_2 & \cdots & A_{m-1} \\ A_{m-1} & A_0 & A_1 & \cdots & A_{m-2} \\ \cdot & \cdot & \ddots & & \cdot \\ \cdot & \cdot & & \ddots & \cdot \\ \cdot & \cdot & & & \cdot \\ A_1 & A_2 & A_3 & \cdots & A_0 \end{bmatrix} \tag{2.1}$$

then we call A a circulant matrix.

The necessary and sufficient condition for A to be a circulant matrix [2]: when

$$k=\begin{cases} v-u & (v\geqslant u) \\ v-u+m & (v<u) \end{cases} \tag{2.2}$$

$A_k = A_{uv}$ of submatrices $A_k = A_{uv}$ (u, v = 0, 1,..., m-1; k = 0, 1,..., m-n) is the square matrices of n x n, and m is the order of partitioned matrix A.

To discuss the characteristics of the circulant matrix we consider

$$\omega=\exp\left[\frac{2\pi}{m}\sqrt{-1}\right]=\exp\left[\frac{2\pi}{m}i\right]$$

which is an n-th root of unity. Let

$$\begin{aligned}\Omega_0^{l\,j}&=\omega^{lj}E_n/\sqrt{m}\\&=\exp\left[\frac{2\pi lj}{m}i\right]E_n/\sqrt{m}. \qquad (l,j=0,1,\cdots,m-1)\end{aligned}$$

where E_n is an n-th order unitary matrix. Clearly, $\Omega_0^{l\,j}$ is also a matrix of order n.

With $\Omega_0^{l\,j}$ as submatrices, we can form an m-th order partitioned matrix

$$\Omega=(\Omega_0^{l,j})$$

We denote by Ω^* the conjugate matrix of Ω, i.e.

$$\Omega^*=\overline{(\Omega_0^{l\,j})}=(\Omega_0^{-l\,j}) \tag{2.3}$$

This is because

$$\begin{aligned}\overline{\Omega_0^{l\,j}}&=\overline{\omega^{lj}}E_n/\sqrt{m}=\exp\left[\frac{-2\pi lj}{m}i\right]E_n/\sqrt{m}\\&=\omega^{-lj}E_n/\sqrt{m}=\Omega_0^{-l\,j}\end{aligned}$$

Below we shall prove a theorem regarding a characteristic of the circulant matrix [2].

Theorem. For a circulant matrix A, we always have

$$\Omega^* A \Omega = C \tag{2.4}$$

where C is a quasidiagonal matrix

$$C = \begin{bmatrix} C_0 & & & & \\ & C_1 & & & \\ & & C_2 & & \\ & & & \ddots & \\ & & & & C_{m-1} \end{bmatrix}$$

$C_l (l = 0, 1, 2, \cdots, m-1)$ is an n-th order square matrix, and

$$C_l = \sqrt{m} \sum_{k=0}^{m-1} \Omega_0^{lk} A_k$$

where A_k is the partitioned matrix A of order m.

Proof. We write each n x n submatrix of the m-th order square matrix Ω^* as $C_{lj} (l, j = 0, 1, 2, \cdots, m-1)$; then

$$C_{lj} = \sum_{u=0}^{m-1} \sum_{v=0}^{m-1} \overline{\Omega_0^{lu}} A_{uv} \Omega_0^{jv} = \sum_{u=0}^{m-1} \sum_{v=0}^{m-1} \Omega_0^{-lu} A_{uv} \Omega_0^{jv}$$

A is a circulant matrix. From the necessary and sufficient condition for A to be a circulant matrix, we have

$$v = \begin{cases} k+u & (v \geq u) \\ k+u-m & (v < u). \end{cases}$$

and

$$A_{uv} = A_k \quad (u, v = 0, 1, \cdots, m-1; \; k = 0, 1, \cdots, m-1)$$

so that

$$C_{lj} = \left(\sum_{u=0}^{m-1} \Omega_0^{(j-l)u} \right) \left(\sum_{k=0}^{m-1} \Omega_0^{jk} A_k \right) \tag{2.5}$$

The first factor in equation (2.5) is

$$\sum_{n=0}^{m-1} \Omega_0^{(j-l)n} = \frac{E_n}{\sqrt{m}} \sum_{n=0}^{m-1} \omega^{(j-l)n}$$

$$= \frac{E_n}{\sqrt{m}} (1+\omega^{j-l}+\omega^{2(j-l)}+\cdots+\omega^{(m-1)(j-l)})$$

$$= \begin{cases} \sqrt{m}E_n & \text{when } j-l=0 \\ 0 & \text{when } 0<j-l\leqslant m-1 \end{cases}$$

The last of the equalities in this equation holds because when $0<j-l\leqslant m-1$,

$$\omega^{j-l}=\exp\left[\frac{2\pi(j-l)}{m}i\right]\neq 1$$

$$\omega^{(j-l)m}=\exp[2\pi(j-l)i]=1$$

and accordingly

$$\frac{E_n}{\sqrt{m}}(1+\omega^{j-l}+\omega^{2(j-l)}+\cdots+\omega^{(m-1)(j-l)})=\frac{E_n}{\sqrt{m}}\frac{(1-\omega^{(j-l)m})}{(1-\omega^{j-l})}=0$$

Therefore, we immediately obtain from equation (2.5)

$$C_{lj}=\sqrt{m}\ E_n\delta_{jl}\sum_{k=0}^{m-1}\Omega_0^{kl}A_k$$

where

$$\delta_{jl}=\begin{cases} 1 & \text{when } j=l \\ 0 & \text{when } j\neq l \end{cases}$$

from which we obtain

$$C_l=\sqrt{m}\ E_n\sum_{k=0}^{m-1}\Omega_0^{kl}A_k \qquad (l=0,1,\cdots,m-1) \tag{2.6}$$

End of proof.

This theorem shows that when an m-th order circulant matrix A is premultiplied by Ω^* and postmultiplied by Ω, it can be converted to a pseudodiagonal matrix $C=\Omega^*A\Omega$. Each submatrix C_l $(l=0,1,\cdots,m-1)$ of C is an n-th order matrix.

We shall use this characteristic of the circulant matrix to derive its inverse.

III. Inversion of the Circulant Matrix

In order to derive the inverse of circulant matrix A, we consider the set of linear algebraic equations with A as coefficient matrix:

$$A\{X\}=\{P\} \tag{3.1}$$

Here A is an m-th order circulant matrix, {X} is an unknown vector, and {P} is a known vector. We assume, of course, that A can be inverted.

In order to use the theorem just proved, we rewrite the above equation. Let

$$\{X\}=\Omega\{Y\}$$

where {Y} is a new, unknown vector, and $\Omega=(\Omega^{ij})$ is the m-th order partitioned matrix described above. Premultiplying both sides of equation (3.1) by Ω^*, we obtain

$$\Omega^* A \Omega\{Y\}=\Omega^*\{P\}$$

From the theorem, the coefficient matrix on the right side of this equation is a pseudodiagonal matrix C, and accordingly

$$C\{Y\}=\Omega^*\{P\}$$

Clearly, C is invertable, and therefore

$$\{Y\}=C^{-1}\Omega^*\{P\}$$

and accordingly $\{X\}=\Omega C^{-1}\Omega^*\{P\}$

This shows that the inverse of circulant matrix A is

$$A^{-1}=\Omega C^{-1}\Omega^* \tag{3.2}$$

For convenience, we write $B=\Omega C^{-1}\Omega^*$. Obviously, B is also an m-th order partitioned matrix, and we write

$$B=(B_{lj}) \quad (l,j=0,1,\cdots,m-1)$$

By the procedure for multiplication of a partitioned matrix, we obtain from equation (3.2)

$$B_{lj}=\sum_{u=0}^{m-1}\sum_{v=0}^{m-1}\Omega_0^{lu}\,C_{uv}^{-1}\,\Omega_0^{-vj}=\sum_{u=0}^{m-1}\Omega_0^{lu}\,C_u^{-1}\,\Omega_0^{-uj}$$

$$=\frac{1}{m}\sum_{u=0}^{m-1}\omega^{lu}\,\omega^{-uj}\,C_u^{-1}=\frac{1}{m}\sum_{u=0}^{m-1}\exp\left[\frac{2\pi(l-j)u}{m}i\right]C_u^{-1}$$

from which it follows that if

$$r=\begin{cases} j-l & (j\geq l) \\ j-l+m & (j<l) \end{cases} \qquad (l, j=0,\ 1,\ \cdots,\ m-1)$$

then $r=0,\ 1,\ \cdots,\ m-1$, and submatrix $B_r=B_{lj}$ is an n-th order square matrix:

$$B_r=\frac{1}{m}\sum_{u=0}^{m-1}\exp\left[\frac{-2\pi ru}{m}i\right]C_u^{-1} \qquad (r=0,\ 1,\ \cdots,\ m-1) \quad (3.3)$$

Finally,

$$A^{-1}=\begin{bmatrix} B_0 & B_1 & B_2 & \cdots & B_{m-1} \\ B_{m-1} & B_0 & B_1 & \cdots & B_{m-2} \\ B_{m-2} & B_{m-1} & B_0 & \cdots & B_{m-3} \\ \cdot & \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ B_1 & B_2 & B_3 & \cdots & B_0 \end{bmatrix} \qquad (3.4)$$

which is the general form for the inverse matrix of circulant matrix A.

The above equation applies for $n\geq 1$.

In particular, when n = 1, we obtain a simple and easily applied result. If n = 1, then circulant matrix A is an m-th order square matrix

$$A=\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{m-1} \\ a_{m-1} & a_0 & a_1 & \cdots & a_{m-2} \\ a_{m-2} & a_{m-1} & a_0 & \cdots & a_{m-3} \\ \cdot & \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix} \qquad (3.5)$$

In this case equation (2.6) is simplified to

$$C_l=\sqrt{m}\sum_{k=0}^{m-1}\frac{1}{\sqrt{m}}\exp\left[\frac{2\pi kl}{m}i\right]a_k=\sum_{k=0}^{m-1}\exp\left[\frac{2\pi kl}{m}i\right]a_k$$
$$(l=0,\ 1,\ \cdots,\ m-1) \qquad (3.6)$$

and equation (3.3) to

$$B_r = \frac{1}{m}\sum_{u=0}^{m-1} \frac{\exp\left[\frac{-2\pi r u}{m} i\right]}{\left(\sum_{k=0}^{m-1} \exp\left[\frac{2\pi u k}{m} i\right] a_k\right)} \tag{3.7}$$

Thus we have obtained an explicit expression for the inverse of matrix A, i.e. $A^{-1} = B$. Accordingly, when $n = 1$, if all of the elements a_0, a_1, $a_2, \ldots, a_{m-1}$ of circulant matrix A are real, then A^{-1} is generally a complex-valued matrix.

IV. An Important Example

Using the conclusion obtained above, we can immediately find the inverse of the circulant matrix

$$A = \begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 4 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 4 & 1 \cdots & 0 & 0 \\ \cdot & & \ddots & \ddots & \ddots & \cdot \\ \cdot & & & \ddots & \ddots & \cdot \\ \cdot & & & & \ddots & \cdot \\ 0 & 0 & \cdots & 1 & 4 & 1 \\ 1 & 0 & \cdots & 0 & 1 & 4 \end{bmatrix}_{m\times m} \tag{4.1}$$

This m-th order circulant matrix is a symmetrical matrix and is frequently encountered [3]. Under the periodicity condition, we obtain this matrix by cubic spline interpolation. As a rule we use equation (4.1) as an algebraic equation for the coefficient matrix when obtaining a solution by the "pursuit" method [1, 3]. But according to the result obtained above, we can immediately write the inverse of equation (4.1), which simplifies the solution process.

Let us find the inverse of matrix (4.1).

Substituting $a_0 = 4$, $a_1 = 1$, $a_{m-1} = 1$, $a_k = 0$ $(k = 2, 3, \ldots, m-2)$ into equation (3.7), we obtain

$$B_r=\frac{1}{m}\sum_{u=0}^{m-1}\frac{\exp\left[-\frac{2\pi ru}{m}i\right]}{4+\exp\left[\frac{2\pi u}{m}i\right]+\exp\left[\frac{2\pi u(m-1)}{m}i\right]}$$

$$=\frac{1}{2m}\sum_{u=0}^{m-1}\frac{\cos\frac{2\pi ur}{m}-i\sin\frac{2\pi ur}{m}}{2+\cos\frac{2\pi u}{m}}$$

$$=\frac{1}{2m}\sum_{u=0}^{m-1}\frac{\cos\frac{2\pi ur}{m}}{2+\cos\frac{2\pi u}{m}}-i\frac{1}{2m}\sum_{u=0}^{m-1}\frac{\sin\frac{2\pi ur}{m}}{2+\cos\frac{2\pi u}{m}}$$

But the sum of all of the terms in the last summation is zero. This is because A is a symmetrical matrix and its inverse must also be symmetrical, so that

$$B_r=B_{m-r}\qquad(r=1,\ 2,\ \cdots,\ m-1)$$

But because

$$\cos\frac{2\pi u(m-r)}{m}=\cos\frac{2\pi ur}{m}$$

$$\sin\frac{2\pi u(m-r)}{m}=-\sin\frac{2\pi ur}{m}$$

it necessarily follows that

$$\sum_{u=0}^{m-1}\frac{\sin\frac{2\pi r}{m}}{2+\cos\frac{2\pi u}{m}}=0$$

and accordingly

$$B_r=\frac{1}{2m}\sum_{u=0}^{m-1}\frac{\cos\frac{2\pi ur}{m}}{2+\cos\frac{2\pi u}{m}}\qquad(r=0,\ 1,\ 2,\ \cdots,\ m-1)\tag{4.2}$$

Therefore we obtain as the inverse of circulant matrix (4.1)

$$A^{-1}=B=\begin{bmatrix} B_0 & B_1 & B_2 & \cdots & B_{m-1}\\ B_{m-1} & B_0 & B_1 & \cdots & B_{m-2}\\ \vdots & \vdots & \ddots & & \vdots\\ & & & \ddots & \\ B_1 & B_2 & B_3 & \cdots & B_0 \end{bmatrix}\tag{4.3}$$

in which the elements B_r (r = 0, 1, 2, ..., m-1) are given by equation (4.2).

In conclusion, we add one further point. In order to test the results obtained above, we consider

$$AB=\begin{bmatrix} 4 & 1 & & & & 1 \\ 1 & 4 & 1 & & & \\ & 1 & 4 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 4 & 1 \\ 1 & & & & 1 & 4 \end{bmatrix} \begin{bmatrix} B_0 & B_1 & B_2 & \cdots & B_{m-2} & B_{m-1} \\ B_{m-1} & B_0 & B_1 & \cdots & B_{m-3} & B_{m-2} \\ B_{m-2} & B_{m-1} & B_0 & \cdots & B_{m-4} & B_{m-3} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ B_2 & B_3 & B_4 & \cdots & B_0 & B_1 \\ B_1 & B_2 & B_3 & \cdots & B_{m-1} & B_0 \end{bmatrix}$$

Any element of product matrix AB can be written

$$4a_k+a_{k+1}+a_{k-1} \qquad (k=0,\ 1,\ \cdots,\ m-1) \tag{4.4}$$

where k = 0 represents the diagonal elements and k = 1, 2, ..., m-1 represents all nondiagonal elements. When k = 0, $a_{k-1} = a_{m-1}$. Because

$$\begin{aligned} 4a_k+a_{k-1}+a_{k+1} &= \frac{1}{2m}\sum_{u=0}^{m-1} \frac{4\cos\frac{2\pi ku}{m}+\cos\frac{2\pi(k-1)u}{m}+\cos\frac{2\pi(k+1)u}{m}}{2+\cos\frac{2\pi u}{m}} \\ &= \frac{1}{2m}\sum_{u=0}^{m-1} \frac{4\cos\frac{2\pi ku}{m}+2\cos\frac{2\pi ku}{m}\cos\frac{2\pi u}{m}}{2+\cos\frac{2\pi u}{m}} \\ &= \frac{1}{m}\sum_{u=0}^{m-1}\cos\frac{2\pi ku}{m} \\ &= \begin{cases} 1 & \text{当}\, k=0 \\ 0 & \text{当}\, k=1,2,\cdots,m-1 \end{cases} \end{aligned}$$

therefore

$$AB=E_m$$

Thus matrix B obtained from equation (4.3) is indeed the inverse of matrix A in equation (4.1).

BIBLIOGRAPHY

1. Ahlberg, J.H.; Nilson, E.N.; and Walsh, J.L. "The Theory of Splines and Their Application," Academic Press, 1967.

2. Wu Jike [2976 7139 0668] and Shao Xiumin. "Circulant Matrices and Their Applications to Structural Computations," JISUAN SHUXUE [COMPUTING MATHEMATICS], Vol 1, No 2, 1979.

3. Feng Kang [7458 1660], et al. "Numerical Computation Methods," National Defense Industry Publishing House, 1978.

8480/6091
CSO: 4008/197

APPLIED SCIENCES

THEORY OF INSTABILITY OF THICK ELECTRON BEAM MASERS

Beijing DIANZI XUEBAO [ACTA ELECTRONICA SINICA] in Chinese Vol 11 No 3, May 83 pp 36-44

[Article by Yang Xianglin [2799 4382 2651] and Wan Suiren [5502 6659 0086] of Nanjing Institute of Technology: "Theory of Instability of Thick Electron Beam Masers"; paper received January 82, finalized May 82.]

[Text] The characteristics of a thick electron beam and its steady-state distribution function are discussed in this paper. The steady-state distribution function of a practical electron beam was expressed in terms of momentum and position variables or all three momentum variables. The complete form of $\nabla\rho f_0$ and the first order perturbed distribution for both H and E modes of the thick electron beam maser were derived and the decoupling nature of the modes in the thick electron beam maser was proven. The general dispersion equations for H and E modes of a cyclotron maser in any mode and cyclotron harmonics at an arbitrary electron beam thickness, including the term $\partial f_0/\partial r_0$ were obtained. Furthermore, the properties of a wave guide maser and a cavity maser were discussed. The results will serve as a theoretical basis for calculating thick electron beam masers. They also are more rigorous and precise calculations for a thin layer beam maser.

I. Zeroth Order Steady-state Distribution Function

Considerable results have been obtained by using the relativistic Vlasov-Maxwell equation to study the instability of betatron masers. However, there are still many problems yet to be investigated. Past studies are mostly limited to the interaction between an electron beam of a specific thickness (twice the Larmor radius) and the H mode in a cylindrical or simplified flat plate system[1-4]. The steady-state distribution of electrons in the phase domain, f_0, is often described by a double function using two momentum δ variables:

$$f_0=\frac{1}{2\pi p_\perp}\delta(p_T-p_{\perp 0})\delta(p_\parallel-p_{\parallel 0}) \qquad (1)$$

where p_T and $p_\parallel$ are the momentum perpendicular and parallel to the uniform steady magnetic field B. From the steady-state Vlasov equation

$$v \cdot \nabla_r f_0 - \frac{e}{c}(v \times B_0) \cdot \nabla_p f_0 = 0 \qquad (2)$$

we know that f_0 is a function of the motion integral ($p_\perp$, and $p_\parallel$) alone only when electrons are uniformly distributed in the position space. In eq.(2), v is the velocity vector of the electron, r is the position variable, p is the momentum variable, e is the charge of the electron, c is the speed of light in vacuum, $\nabla r f_0$ is the gradient of f_0 in the position space, and $\nabla p f_0$ is the gradient of f_0 in the momentum space. Therefore, the distribution of electrons in position space described by eq.(1) is uniform and uncertain. It cannot reflect the position of the electron beam in the wave guide. There is some deviation from reality. The dispersion equation and interaction power thus obtained must contain some error.

In practice, electrons in a cyclotron are not uniformly distributed in position, i.e. $\nabla r f_0 \neq 0$. f_0 must be a function of (r,p). The center of an electron guide cannot possibly be located on a cylindrical surface to form an electron beam with a thickness twice the Larmor radius. Let us assume that this guide center deviates from the ideal cylinder by ($\pm\nabla R_0$) and the distribution is uniform in this region, as shown in the thick beam model in Figure 1. Jeans theorem points out that any function created by the integral of motion will satisfy the steady-state Vlasov equation. Therefore, the function f_0 can be created by the integral of motion from $p_\perp$, $p_\parallel$ and r_0 (electron guide center radius) or $p_\perp$, $p_\parallel$ and p_0 (sine component of the angular momentum of the electron). The f_0 formed by the integral of motion (r_0, $p_\perp$, $p_\parallel$) to describe the model in Figure 1 is

$$f_0(r_0, p_\perp, p_\parallel) = N_z \frac{H[\Delta R_0^2 - (r_0 - R_0)^2]}{4\pi R_0 \Delta R_0} g(p_\perp, p_\parallel) \qquad (3)$$

where $g(p_\perp, p_\parallel) = \frac{1}{2\pi p_\perp}\delta(p_\perp - p_{\perp 0})\delta(p_\parallel - p_{\parallel 0})$, N_z is the linear density of the electron in the z-direction, $r_0^2 = r^2 + r_L^2 - 2rr_L \sin\hat{\phi}$, $r_L = p_\perp / m_0\omega_c$, $\omega_c = eB_0/m_0c$,

m_0 is the rest mass of electron, $\hat{\phi} = \phi - \theta$, H(x) represents that all the electron guide center r_0 is in the [$R_0 - \Delta R_0$, $R_0 + \Delta R_0$] range. When $x \geqslant 0$ and $x < 0$, H(x) is 1 and 0, respectively. When $\Delta R_0 \to 0$, it corresponds to an electron ring whose thickness is the diameter of the Larmor center and its guide center is located at R_0. In this case

$$\lim_{\Delta R_0 \to 0} \frac{H[\Delta R_0^2 - (r_0 - R_0)^2]}{4R_0 \Delta R_0} = \delta(r_0^2 - R_0^2) \qquad (4)$$

The relation among vectors r_o, r and r_L are shown in Figure 2.

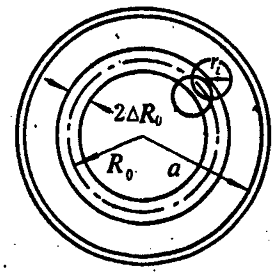


Figure 1. Cross-section of Wave Guide and Electron Beam Model. (B_O and z axes are pointing out of the paper, a is the wave guide radius, R_O is the radius of the guide center and r_L is the Larmor radius.)

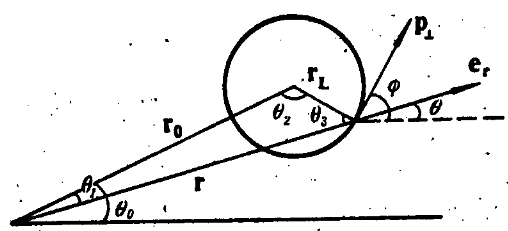


Figure 2. Relation Among r_o, r and r_L.

The electron beam density function in the wave guide is $n(r)=\int d^3p(r,p)$, when $\Delta R_O \geq r_{Lo}$

$$n(r)=\frac{N_z}{4\pi R_0 \Delta R_0}\begin{cases} \frac{1}{2}-\frac{\phi_{10}}{\pi} & R_0+\Delta R_0-r_{L0}\leqslant r\leqslant R_0+\Delta R_0+r_{L0} \\ 1 & R_0-\Delta R_0+r_{L0}<r<R_0+\Delta R_0-r_{L0} \\ \frac{1}{2}+\frac{\phi_{20}}{\pi} & R_0-\Delta R_0-r_{L0}\leqslant r\leqslant R_0-\Delta R_0+r_{L0} \\ 0 & \text{in other regions} \end{cases} \qquad (5)$$

When $\Delta R_0 < r_{L0}$

$$n(r)=\frac{N_s}{4\pi R_0 \Delta R_0}\begin{cases}\dfrac{1}{2}-\dfrac{\phi_{10}}{\pi} & R_0-\Delta R_0+r_{L0}\leqslant r\leqslant R_0+\Delta R_0+r_{L0}\\ \dfrac{\phi_{20}-\phi_{10}}{\pi} & R_0+\Delta R_0-r_{L0}<r<R_0-\Delta R_0+r_{L0}\\ \dfrac{1}{2}+\dfrac{\phi_{20}}{\pi} & R_0-\Delta R_0-r_{L0}\leqslant r\leqslant R_0+\Delta R_0-r_{L0}\\ 0 & \text{in other regions}\end{cases} \tag{6}$$

When $\Delta R_0=0$

$$n(r)=\frac{N_s H(R_0+r_{L0}-r)\cdot H(r-R_0+r_{L0})}{\pi^2[(R_0+r_{L0})^2-r^2]^{1/2}[r^2-(R_0-r_{L0})^2]^{1/2}} \tag{7}$$

where $$r_{L0}=\frac{p_{\perp 0}}{m_0\omega_c};\ \phi_{10}=\sin^{-1}\frac{r^2+r_{L0}^2-(R_0+\Delta R_0)^2}{2rr_{L0}};\ \phi_{20}=\sin^{-1}\frac{r^2+r_{L0}^2-(R_0-\Delta R_0)^2}{2rr_{L0}}.$$

The gradient of the steady-state function in momentum space as defined in eq. (3) is

$$\nabla_p f_0(r_0,p_\perp,p_\parallel)=\mathbf{e}_\perp\left[\frac{\partial f_0}{\partial p_\perp}+\frac{r_L}{r_0 m_0\omega_c}\frac{\partial f_0}{\partial r_0}\right]-\mathbf{e}_\theta\frac{r}{r_0 m_0\omega_c}\frac{\partial f_0}{\partial r_0}+\mathbf{e}_\parallel\frac{\partial f_0}{\partial p_\parallel} \tag{8}$$

where $\dfrac{r_L}{r_0 m_0\omega_c}\dfrac{\partial f_0}{\partial r_0}$ and $\dfrac{r}{r_0 m_0\omega_c}\dfrac{\partial f_0}{\partial r_0}$ are two terms due to the introduction of the integral of motion r_0 to the steady-state distribution. They have not been included in the literature.

II. First Order Perturbed Distribution Function

Let us assume that the electron beam is dilute enough to neglect the space charge field. Based on the model shown in Figure 1, a cavity cyclotron maser is formed by a circular wave guide with a node L. The operating mode in the cavity is H_{mng} and the standing wave field in the cavity can be decomposed as the superposition of forward and reverse waves.

$$E=E^{+}+E^{-} \tag{9}$$

The field can be expanded at the guide center as follows:

$$
\left.
\begin{aligned}
E_{\perp}^{\pm} &= \pm \frac{E_0}{2} \sum_{l=-\infty}^{\infty} J'_{-l}(k'_n r_L) J_{m-l}(k'_n r_0) G^{\pm}_{m,l}(\theta,\phi,z,t) \\
E_{\phi}^{\pm} &= \pm j \frac{E_0}{2} \sum_{l=-\infty}^{\infty} \frac{l}{k'_n r_L} J_{-l}(k'_n r_L) J_{m-l}(k'_n r_0) G^{\pm}_{m,l}(\theta,\phi,z,t) \\
B_{z}^{\pm} &= \pm j \frac{E_0}{2} \frac{k'_n c}{\omega} \sum_{l=-\infty}^{\infty} J_{-l}(k'_n r_L) J_{m-l}(k'_n r_0) G^{\pm}_{m,l}(\theta,\phi,z,t) \\
B_{\perp}^{\pm} &= -j \frac{E_0}{2} \frac{k_{\parallel} c}{\omega} \sum_{l=-\infty}^{\infty} \frac{l}{k'_n r_L} J_{-l}(k'_n r_L) J_{m-l}(k'_n r_0) G^{\pm}_{m,l}(\theta,\phi,z,t) \\
B_{\phi}^{\pm} &= \frac{E_0}{2} \frac{k_{\parallel} c}{\omega} \sum_{l=-\infty}^{\infty} J_{-l}(k'_n r_L) J_{m-l}(k'_n r_0) G^{\pm}_{m,l}(\theta,\phi,z,t)
\end{aligned}
\right\} \tag{10}
$$

where $k_{\parallel}=q\pi/L$ and q is a positive integer, $k'_n=\nu'_{mn}/a$ and ν'_{mn} is the nth non-zero root of $J'_m(x)=0$, $G^{\pm}_{m,l}=\exp\{j[m\theta_0+l\theta_2+k_{\parallel}z-\omega t]\}$ and $\theta_0=\theta+\theta_1$ (see Figure 2). The superscript "+" is used for a wave guide cyclotron maser.

Similarly, the expansion of the field in the E mode near the wave guide center can be obtained.

The first order perturbed distribution function is:

$$f_0(r,p,t)=e\int_{t_1}^{t} dt'\left[E'+\frac{1}{c}v'\times B'\right]\cdot\nabla_p f_0(r',p') \tag{11}$$

where all terms with "'" are functions of t'. For a cavity maser, $t_1=t-z/v_{\parallel}$. For a wave guide maser, $t_1\to-\infty$.

Among all the terms in the expansion near the center of the wave guide, only $G_{m,l}(\theta,\phi,z,t,)$ is a function of t. From the unperturbed equation of motion we get

$$G_{m,l}(\theta',\phi',z',t')=G_{m,l}(\theta,\phi,z,t)\exp[j(\omega-k_{\parallel}v_{\parallel}-l\Omega)\tau],\ \Omega=\omega_c/\gamma,\ \gamma=\left(1-\frac{v^2}{c^2}\right)^{-1/2},\ \tau=t-t'.$$

By substituting eq. (8) and (10) into (11), we get

$$f_{1H}^{\pm}=\pm j\frac{eE_0}{2}\sum_{l=-\infty}^{\infty}\left\{\left[\left(1\mp\frac{k_{\parallel}v_{\parallel}}{\omega}\right)\frac{\partial f_0}{\partial p_{\perp}}\pm\frac{k_{\parallel}v_{\parallel}}{\omega}\frac{\partial f_0}{\partial p_{\parallel}}\right]J'_{-l}(k'_n r_L)J_{m-l}(k'_n r_0)\right.$$
$$\left.-\frac{\partial f_0/\partial r_0}{m_0\omega_c}\left(J_{-l}(k'_n r_L)J'_{m-l}(k'_n r_0)\,e^{\pm}+\frac{m-l}{r_0}r_L J'_{-l}(k'_n r_L)J_{m-l}(k'_n r_0)\Omega\right)\right]\right\}\frac{X^{\pm}}{e^{\pm}}G_{m,l} \tag{12}$$

where $e^{\pm}=\omega\mp k_{\parallel}v_{\parallel}-l\Omega$, $\mathrm{Im}(e^{\pm})>0$, $X^{\pm}=1-\exp(je^{\pm}z/v_{\parallel})$, and "±" represents the perturbation caused by the forward and reverse wave field, respectively. In a wave guide maser, only the "+" terms are considered. Moreover, $\lim_{X\to\infty}X^{t}=1$. The subscript "H" represents the H mode perturbed distribution function.

Similarly, we can get the E mode perturbed distribution function

$$f_{1E}=\frac{eE_0\omega}{k_{\mu}c}\sum_{l=-\infty}^{\infty}\left\{\left[\frac{l}{k_{\mu}r_L}\left(\frac{k_{\parallel}c}{\omega}-\frac{v_{\parallel}}{c}\right)\frac{\partial f_0}{\partial p_{\perp}}+\left(\frac{l\Omega}{k_{\mu}c}-\frac{k_{\mu}c}{\omega}\right)\frac{\partial f_0}{\partial p_{\parallel}}\right]\right.$$
$$\left.+\frac{\partial f_0/\partial r_0}{m_0\omega_c}\frac{m-l}{k_{\mu}r_0}\left(\frac{v_{\parallel}}{c}-\frac{k_{\parallel}c}{\omega}\right)\right\}\frac{J_{-l}(k_{\mu}r_L)J_{m-l}(k_{\mu}r_0)}{\omega-k_{\parallel}v_{\parallel}-l\Omega}G_{m,l} \tag{13}$$

where $K_{\mu}=\nu_{m\mu}/a$ and $\nu_{m\mu}$ is the μth non-zero root of $J_m(x)=0$

III. Decoupling Nature of Modes

The dispersion equation, mode coupling and instability can be studied by using the perturbed distribution function.

Let us assume that various H and E modes exist in the wave guide. The total B_z field of the H modes is $B_z=\sum_i B_z^i$ where i is the H mode number. The total E_z field of the E modes is $E_z=\sum_s E_z^s$ where s is the E mode sequence. Using Maxwell equation, we get

$$\sum_i\left(\frac{\omega_i^2}{c^2}-k_{ni}^2-k_{ci}'^2\right)B_z^i=-\frac{4\pi}{c}\sum_i(\nabla\times J_i)_z-\frac{4\pi}{c}\sum_s(\nabla\times J_s)_z \tag{14}$$

$$\sum_s\left(\frac{\omega_s^2}{c}-k_{\parallel s}^2-k_{cs}'^2\right)E_z=4\pi j\sum_i k_{\parallel i}\left(\rho_i-\frac{\omega_i}{k_{\parallel i}c^2}J_{iz}\right)+4\pi j\sum_s k_{\parallel s}\left(\rho_s-\frac{\omega_s}{k_{\parallel s}c^2}J_{sz}\right) \tag{15}$$

where k'_{ci} and k'_{cs} are the cutoff wave number of the H_i and E_s modes, respectively. J_i, J_s and ρ_i, ρ_s are the first order and zeroth order moment of f^i_{1H} and f^i_{1E} with respect to velocity v, respectively. f^i_{1H} and f^i_{1E} are the perturbed distribution functions of parallel H^i and E^s modes, respectively.

Let us multiply both sides of eq. (14) by B_Z^{m*} and both sides of eq. (15) by E_Z^{n*}. Then, we get the following by integrating in the wave guide:

$$\sum_i\left(\frac{\omega_i^2}{c^2}-k_{\parallel i}^2-k_{ci}'^2\right)\lim_{z\to\infty}\int_0^Z dz\int_0^{2\pi}d\theta\int_0^a r\,dr\,B_z^i\cdot B_z^{m*}$$
$$=-\frac{4\pi}{c}\sum_i\lim_{Z\to\infty}\int_0^Z dz\int_0^{2\pi}d\theta\int_0^a r\,dr\,B_z^{m*}\cdot(\nabla\times J_i)_z-\frac{4\pi}{c}\lim_{Z\to\infty}\int_0^z dz$$
$$\int_0^{2\pi}d\theta\int_0^a r\,dr\,B_z^{m*}(\nabla\times J_s)_z \tag{16}$$

$$\sum_{s\,i}\left(\frac{\omega_i^2}{c^2}-k_{\parallel s}^2-k_{ci}'^2\right)\lim_{Z\to\infty}\int_0^Z dz\int_0^{2\pi}d\theta\int_0^a r\,dr\,E_z^i\cdot E_z^{n*}$$
$$=4\pi j\sum_i\lim_{Z\to\infty}\int_0^Z dz\int_0^{2\pi}d\theta\int_0^a r\,dr\,k_{\parallel i}\left(\rho_i-\frac{\omega_i}{k_{\parallel i}c^2}J_{iz}\right)\cdot E_z^{n*}$$
$$+4\pi j\sum_s\lim_{Z\to\infty}\int_0^Z dz\int_0^{2\pi}d\theta\int_0^a r\,dr\,k_{\parallel s}\left(\rho_s-\frac{\omega_s}{k_{\parallel s}c^2}J_{sz}\right)\cdot E_z^{n*} \tag{17}$$

For a sufficiently dilute electron beam, the function to be integrated on the left of eq. (16) is the product of orthogonal modes. All i≠m terms are zero. Then, eq. (16) becomes:

$$\left(\frac{\omega_m^2}{c^2}-k_{\parallel m}^2-k_{cm}'^2\right)\lim_{Z\to\infty}\int_0^Z dz\int_0^{2\pi}d\theta\int_0^a r\,dr\,B_z^m\cdot B_z^{m*}$$
$$=\left(\frac{\omega_m^2}{c^2}-k_{\parallel m}^2-k_{cm}'^2\right)\lim_{Z\to\infty}Z\cdot SN_{mn}\frac{k_{cm}'^2c^2}{\omega_m^2} \tag{18}$$

where $$S=\pi a^2;\quad N_{mn}=\left(1-\frac{m^2}{k_n^2a^2}\right)J_m^2(k_n'a),\,k_n'=k_{cmo}'$$

By dividing both sides of eq. (16) by $\lim_{Z\to\infty}Z\cdot SN_{mn}\frac{k_{cm}'c^2}{\omega_m^2}$, we get

$$
\begin{aligned}
\frac{\omega_m^2}{c^2}-k_{\parallel m}^2-k_{cm}'^2 &= \frac{e^2\pi\omega_m}{m_0 k_{cm}' c^2 S N_{mn}} \sum_i \sum_{l=-\infty}^{\infty} \sum_{l'=-\infty}^{\infty} \int_0^{\infty} dp_\perp \int_{-\infty}^{\infty} dp_\parallel \int_0^{a} r_0 dr_0 \\
&\cdot k_{ci}' E_0^m E_0^i \frac{p_\perp^i}{\gamma} J_{-l'}'(k_{cm}' r_L) J_{m-l'}(k_{cm}' r_0) \left\{\left[\left(1-\frac{k_{\parallel i} v_\parallel}{\omega_i}\right)\frac{\partial f_0}{\partial p_\perp}\right.\right. \\
&\left. + \frac{k_{\parallel i} v_\parallel}{\omega_i}\frac{\partial f_0}{\partial p_\parallel}\right] J_{-l}'(k_{ci}' r_L) J_{i-l}(k_{ci}' r_0) - \frac{\partial f_0/\partial r_0}{m_0\omega_c}\Big[J_{-l}(k_{ci}' r_L) \\
&\left.\left. \cdot J_{i-l}(k_{ci}' r_0)\epsilon^+ + \frac{i-l}{r_0} r_L J_{-l}'(k_{ci}' r_L) J_{i-l}(k_{ci}' r_0)\Omega\right]\right\}\frac{1}{\epsilon^+} \\
&\cdot \lim_{Z\to\infty}\frac{1}{Z}\int_0^Z dz \int_0^{2\pi} d\theta_0 \int_0^{2\pi} d\theta_2 \cdot \exp\{j[(i-m)\theta_0+(l-l')\theta_2 \\
&+ (k_{\parallel i}-k_{\parallel m})z-(\omega_i-\omega_m)t]\} - j\frac{2\pi e^2\omega_m}{m_0 k_{cm}' c^3 S N_{mn}} \sum_s \sum_{l=-\infty}^{\infty} \\
&\cdot \sum_{l'=-\infty}^{\infty} \int_0^{\infty} dp_\perp \int_{-\infty}^{\infty} dp_\parallel \int_0^a r_0 dr_0 \cdot \omega_s E_0^m \cdot E_0^s \frac{p_\perp^2}{\gamma} J_{-l'}'(k_{cm}' r_L) \\
&\cdot J_{m-l'}(k_{cm}' r_0)\left\{\left[\frac{l}{k_{cs} r_L}\left(\frac{k_{\parallel s} c}{\omega_s}-\frac{v_\parallel}{c}\right)\frac{\partial f_0}{\partial p_\perp}+\left(\frac{l\Omega}{k_{cs} c}-\frac{k_{cs} c}{\omega_s}\right)\right.\right. \\
&\left.\left.\cdot \frac{\partial f_0}{\partial p_\parallel}\right] + \frac{\partial f_0/\partial r_0}{m_0\omega_c}\frac{s-l}{k_{cs} r_0}\left(\frac{v_\parallel}{c}-\frac{k_{\parallel s} c}{\omega_s}\right)\right\}\frac{J_{-l}(k_{cs} r_L) J_{s-l}(k_{cs} r_0)}{\omega_s - k_{\parallel s} v_\parallel - l\Omega} \\
&\cdot \lim_{Z\to\infty}\frac{1}{Z}\int_0^Z dz \int_0^{2\pi} d\theta_0 \int_0^{2\pi} d\theta_2 \cdot \exp\{j[(s-m)\theta_0+(l-l')\theta_2 \\
&+ (k_{\parallel s}-k_{\parallel m})z-(\omega_s-\omega_m)t]\} \qquad (19)
\end{aligned}
$$

where E_0^m, E_0^i and E_0^s are the amplitudes of the H_m, H_i and H_s modes. The integration under the summation sign in eq. (19) is not zero only when i=m, l=l' and $k_{\parallel i}=k_{\parallel m}$. This suggests that the H_m and H_i modes ($i\neq m$) are not coupled through the electron beam. Similarly, the integral under $\sum_s$ is not zero only when s=m, l=l' and $k_{\parallel s}=k_{\parallel m}$. However, because $k_{\parallel s}$ is the propagation constant for the E mode and $k_{\parallel m}$ is that for H mode, they will never be equal unless it is a degenerate mode. If it is degenerate, then $s\neq m$. Therefore, this integral is always zero, which means that the H_m mode will not couple with the E_s mode through the electron beam. Similar conclusions can be obtained for the integral in eq. (17) based on the principle of symmetry.

We can see that the E and H modes do not couple in a dilute electron cyclotron maser. They are completely decoupled, even in a degenerated mode. The energy exchange between a mode and the electron beam is also independent in a cavity maser. This decoupling nature is not related to the zeroth order distribution function, i.e. unrelated to the equilibrium state of the electrons. This is in agreement with the anticipated result when we use linear approximation and neglect the space charge field. Hence, it can be used to study the dispersion equation, instability and power function for each individual mode.

IV. Dispersion Equation

The dispersion equation for the H mode of a circular wave guide-thick electron beam model is:

$$\frac{\omega^2}{c^2}-k_{\parallel}^2-k_n'^2=\frac{4\pi\nu}{\gamma_0 S N_{mn}'}\left\{\varepsilon_{H1}(a_0,a_L,\Delta a_0)+\frac{\Omega_0\varepsilon_{H2}(a_0,a_L,\Delta a_0)}{\omega-k_{\parallel}v_{\parallel 0}-l\Omega_0}\right.$$
$$\left.+\left[\frac{(\omega-k_{\parallel}v_{\parallel 0})Q^H_{m,l}(a_0,a_L)}{\omega-k_{\parallel}v_{\parallel 0}-l\Omega_0}-\frac{\beta_{\perp 0}^2(\omega^2-k_{\parallel}^2c^2)H^H_{m,l}(a_0,a_L)}{(\omega-k_{\parallel}v_{\parallel 0}-l\Omega_0)^2}\right]\cdot\zeta(a_0,\Delta a_0)\right\} \tag{20}$$

where the subscripts i and m are dropped in ω, $k_{\parallel}$ and k_n'.

$$\nu=e^2N_z/m_0C^2;\quad a_0=k_n'R_0;$$

$$a_L=k_n'r_{L0};\quad \Delta a_0=k_n'\Delta R_0;\quad \gamma_0=\left(1+\frac{p_{\perp 0}^2+p_{\parallel 0}^2}{m_0^2c^2}\right)^{1/2};\quad \Omega_0=\frac{\omega_c}{\gamma_0};\quad \beta_{\perp 0}=\frac{p_{\perp 0}}{m_0\gamma_0 c};$$

$$H^H_{m,l}(a_0,a_L)=[J_l'(a_L)J_{m-l}(a_0)]^2 \tag{21}$$

$$Q^H_{m,l}(a_0,a_L)=a_L\left(\frac{l^2}{a_L^2}-1\right)[J_l^2(a_L)]'J_{m-l}^2(a_0) \tag{22}$$

$$\zeta(a_0,\Delta a_0)=\frac{J_{m-l}^{-2}(a_0)}{4a_0\Delta a_0}\{a_{02}^2[J_{m-l}^2(a_{02})-J_{m-l-1}(a_{02})J_{m-l+1}(a_{02})]$$
$$-a_{01}^2[J_{m-l}^2(a_{01})-J_{m-l-1}(a_{01})J_{m-l+1}(a_{01})]\} \tag{23}$$

$$\left.\begin{aligned}\varepsilon_{H1}(a_0,a_L\Delta a_0)&=\sum_{l=-\infty}^{\infty}\frac{a_L}{4a_0\Delta a_0}[J_l^2(a_L)]'\{a_{01}[J_{m-l}^2(a_{01})]'-a_{02}[J_{m-l}^2(a_{02})]'\}\\ \varepsilon_{H2}(a_0,a_L,\Delta a_0)&=\frac{a_L^2}{\Delta a_0}\,\frac{m-l}{a_0}J_l'^2(a_L)[J_{m-l}^2(a_{02})-J_{m-l}^2(a_{01})]\end{aligned}\right\} \tag{24}$$

$$a_{01}=a_0-\Delta a_0,\quad a_{02}=a_0+\Delta a_0$$

The E mode dispersion equation is

$$\frac{\omega^2}{c^2}-k_{\parallel}^2-k_{\mu}^2=\frac{4\pi\nu}{\gamma_0 S K_{mu}}\left\{\left[\frac{Q^E_{ml}(b_0,b_L)\zeta(b_0,\Delta b_0)+\varepsilon_E(b_0,b_L,\Delta b_0)}{\Omega_0(\omega-k_n v_{\parallel 0}-l\Omega_0)}\right.\right.$$
$$\left.-\frac{H^E_{m,l}(b_0,b_L)\cdot\zeta(b_0,\Delta b_0)}{(\omega-k_{\parallel}v_{\parallel 0}-l\Omega_0)^2}\right]\left(\frac{\omega v_{\parallel 0}}{c}-k_{\parallel}c\right)^2$$
$$\left.-\frac{\left(\frac{\omega^2}{c^2}-k_{\parallel}^2-k_{\mu}^2\right)\frac{l\Omega_0}{k_{\mu}^2}\cdot H^E_{m,l}(b_0,b_L)\zeta(b_0,\Delta b_0)}{\omega-k_{\parallel}v_{\parallel 0}-l\Omega_0}\right\} \tag{25}$$

where

$$b_0=k_\mu R_0;\quad b_L=k_\mu r_{L0};\quad \Delta b_0=k_\mu \Delta R_0;\quad K_{m\mu}=J_m'^2(k_\mu a);$$

$$H^E_{m,l}(b_0,b_L)=J_l^2(b_L)J^2_{m-l}(b_0) \tag{26}$$

$$Q^E_{m,l}(b_0,b_L)=\frac{l}{b_L}[J_l^2(b_L)]'J^2_{m-l}(b_0) \tag{27}$$

$$\varepsilon_E(b_0,b_L,\Delta b_0)=\frac{m-l}{b_0}J_l^2(b_L)\frac{1}{2\Delta b_0}[J^2_{m-l}(b_{02})-J^2_{m-l}(b_{01})] \tag{28}$$

In dispersion equations (20) and (25), ζ is the geometric correction factor for a thick electron beam relative to a thick electron beam with a thickness of twice the radius of the Larmor radius. $\zeta\leqslant 1$ and $\lim_{\Delta R_0\to 0}=1$. ε_H and ε_E are added due to the introduction of the motion integral r_0 and the $\partial f_0/\partial r_0$ term.

When $\Delta R_0\to 0$, the dispersion equation for a thick electron beam with a thickness of twice the Larmor radius can be obtained from eq. (20) and (25). In this case, we have

$$\left.\begin{aligned}
\varepsilon_{H1}(a_0,a_L,0)&=-\sum_{l=-\infty}^{\infty}\frac{a_L}{2}[J_l^2(a_l)]'\left\{\frac{1}{a_0}[J^2_{m-l}(a_0)]'+[J^2_{m-l}(a_0)]''\right\}\\
\varepsilon_{H2}(a_0,a_L,0)&=2a_L^2\frac{m-l}{a_0}J_l'^2(a_L)[J^2_{m-l}(a_0)]'
\end{aligned}\right\} \tag{29}$$

$$\varepsilon_E(b_0,b_L,0)=\frac{m-l}{b_0}J_l^2(b_L)[J^2_{m-l}(b_0)]' \tag{30}$$

When $\Delta R_0=R_0$, $a_{01}=b_{01}=0$. In this situation, the inner edge of the thick electron wave guide center circle shrinks back to r=0. The electron beam becomes solid. From ea. (20) and (25), we can obtain a solid dispersion equation. For the H mode, we have:

$$\frac{\omega^2}{c^2}-k_\parallel^2-k_n'^2=\frac{4\pi\nu}{\gamma_0 S N_{mn}}\left\{\varepsilon^s_{H1}(a_0,a_L)+\frac{\Omega_0\varepsilon^s_{H2}(a_0,a_L)+(\omega-k_\parallel v_{\parallel 0})Q^{HS}_{m,l}(a_0,a_L)}{\omega-k_\parallel v_{\parallel 0}-l\Omega_0}\right.$$
$$\left.-\frac{\beta^2_{\perp 0}(\omega^2-k_\parallel^2c^2)H^{HS}_{m,l}(a_0,a_L)}{(\omega-k_\parallel v_{\parallel 0}-l\Omega_0)^2}\right\} \tag{31}$$

$$\left.\begin{aligned}
\varepsilon^s_{H1}(a_0,a_L)&=-\sum_{l=-\infty}^{\infty}\frac{a_L}{a_0}[J_l^2(a_L)]'[J^2_{m-l}(a_0)]'\\
\varepsilon^s_{H2}(a_0,a_L)&=\frac{4a_L^2}{a_0^2}(m-l)J_l'^2(a_L)J^2_{m-l}(a_0)
\end{aligned}\right\} \tag{32}$$

$$H^{HS}_{m,l}(a_0,a_L)=J_l'^2(a_L)[J^2_{m-l}(a_0)-J_{m-l-1}(a_0)J_{m-l+1}(a_0)] \tag{33}$$

$$Q^{HS}_{m,l}(a_0,a_L)=a_L\left(\frac{l^2}{a_L^2}-1\right)[J_l^2(a_L)]'[J^2_{m-l}(a_0)-J_{m-l-1}(a_0)J_{m-l+1}(a_0)] \tag{34}$$

where a_0 is the product of the inner radius of the electron guide center circle and k_n'.

With respect to the E mode

$$\frac{\omega^2}{c^2}-k_\parallel^2-k_\mu^2=\frac{4\pi\nu}{\gamma_0 S K_{m\mu}}\left\{\left[\frac{\theta_{m,l}^{ES}(b_0,b_L)+\varepsilon_E^S(b_0,b_L)}{\Omega_0(\omega-k_z v_{\parallel 0}-l\Omega_0)}-\frac{H_{m,l}^{ES}(b_0 b_L)}{(\omega-k_\parallel v_{\parallel 0}-l\Omega_0)^2}\right]\cdot\left(\frac{\omega v_{\parallel 0}}{c}-k_\parallel c^2\right)-\frac{\left(\frac{\omega^2}{c^2}-k_\parallel^2-k_\mu^2\right)\frac{l\Omega_0}{k_\mu^2}H_{m,l}^{ES}(b_0,b_L)}{\omega-k_\parallel v_{\parallel 0}-l\Omega_0}\right\} \tag{35}$$

$$H_{m,l}^{ES}(b_0,b_L)=J_l^2(b_L)[J_{m-l}^2(b_0)-J_{m-l-1}(b_0)J_{m-l+1}(b_0)] \tag{36}$$

$$Q_{m,l}^{ES}(b_0,b_L)=\frac{l}{b_L}[J_l^2(b_L)]'[J_{m-l}^2(b_0)-J_{m-l-1}(b_0)J_{m-l+1}(b_0)] \tag{37}$$

$$\varepsilon_E^S(b_0,b_L)=\frac{2(m-l)}{b_0^2}J_l^2(b_L)J_{m-l}^2(b_0) \tag{38}$$

where b_0 is the product of the outer radius of the electron guide center circle and $k\mu$.

V. Calculation of Characteristics of Thick Electron Beam Masers

1. Deviation from Linearity and Rate of Increase of Wave Guide Masers.

Let $\omega=\omega_0+\delta\omega$, and $\omega_0=c\sqrt{k_\parallel^2+k_n'^2}=k_\parallel v_{\parallel 0}+l\Omega_0$, then eq. (20) becomes

$$\delta^3\omega-\frac{2\pi\nu c^2}{\gamma_0 S N_{mn}\omega_0}\{\varepsilon_{H1}\delta^2\omega+\Omega_0(\varepsilon_{H2}+lQ_{m,l}^H\zeta)\delta\omega-\beta_{\perp 0}^2 c^2 k_n'^2 H_{m,l}^H\zeta=0 \tag{39}$$

We can then calculate the linearity deviation, $\delta\omega_r$, and rate of increase, $\delta\omega_i$, are

$$\delta\omega_r\approx\delta\omega_r^0[\zeta(a_0,\Delta a_0)]^{1/3},\qquad \delta\omega_r^0=\frac{1}{2}\left[\frac{2\pi\nu k_n'^2 c^2 v_{\perp 0}^2 H_{m,l}^H}{\gamma_0 S N_{mn}\omega_0}\right]^{1/3} \tag{40}$$

$$\delta\omega_i=\sqrt{3}\,\delta\omega_r \tag{41}$$

where $\delta\omega_r^0$ is the frequency deviation of an ideal single layer beam, $[\zeta(a_0,\Delta a_0)]^{1/3}$ is a correction for the deviation of the center of the guide.

2. Power of Interaction in Cavity Masers.

The power of interaction between the electron beam and the field is

$$P_e=\frac{1}{2}\int_v E\cdot J^* dV=P_{er}+jP_{ei} \tag{42}$$

$$P_{er}=\frac{\nu E_0^2c^2}{4\omega\gamma_0}\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\tau_0\sin^2\frac{\Delta}{2}\Big[k^H_{m,l}\ \tau_0\beta^2_{\perp 0}\Big\{(\omega^2-k^2_\parallel c^2)\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\cdot\Big[\frac{1}{2}\cos\frac{\Delta}{2} -\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)\Big]L-k_\parallel C^2\Big[\cos\frac{\Delta}{2}-\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)\Big]\Big\}+q^H_{m,l}\ L\cdot\left(\frac{\omega-k_\parallel v_{\parallel 0}}{\Delta}-\frac{\omega+k_\parallel v_{\parallel 0}}{\Delta'}\right) -\varepsilon_H\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)L\Omega_0\Big] \tag{43}$$

$$P_{ei}=\frac{\nu E_0^2c^2}{8\omega\gamma_0}\tau_0\Big[h^H_{m,l}\ \tau_0\beta^2_{\perp 0}\Big\{(\omega^2-k^2_\parallel c^2)L\Big[\left(\cos\Delta-2\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)\sin\Delta\right) \cdot\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)^2+\left(\frac{1}{\Delta^2}+\frac{1}{\Delta'^2}\right)\Big]+2k_\parallel c^2\Big[\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)\sin\Delta-\cos\Delta\Big]\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\Big\} -q^H_{m,l}L\Big\{\frac{\omega-k_\parallel v_{\parallel 0}}{\Delta}\Big[1-\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\sin\Delta\Big]+\frac{\omega+k_\parallel v_{\parallel 0}}{\Delta'}\Big[1+\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\sin\Delta\Big]\Big\} +\varepsilon_H L\Omega_0\Big[\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)-\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)^2\sin\Delta\Big]\Big\} \tag{44}$$

where $h^H_{m,l}=H^H_{m,l}(a_0,a_L)\zeta(a_0,\Delta a_0)$; $q^H_{m,l}=Q^H_{m,l}(a_0,a_L)\cdot\zeta(a_0,\Delta a_0)+\varepsilon^l_{H1}(a_0,a_L,\Delta a_0)$; $\varepsilon_H= l\varepsilon^l_{H1}(a_0,a_L,\Delta a_0)-\varepsilon_{H2}(a_0,a_L,\Delta a_0)$; $\tau_0=L/v_{\parallel 0}$; $\Delta=(\omega-k_\parallel v_{\parallel 0}-l\Omega_0)\tau_0$; $\Delta'=(\omega+k_\parallel v_{\parallel 0})-l\Omega_0)\tau_0$;

ε^l_{H1} is the l term in ε_{H1}.

3. Resonance Current and Frequency Offset.

When a cavity maser works stably

$$-P_{er}=\omega W/Q_L \tag{45}$$

where Q is the carrier quality factor in the cavity. The energy stored in the cavity is
$$W=\frac{E_0^2}{16}La^2\left(1-\frac{m^2}{k_n'^2a^2}\right)J_m^2(k_n'a)。$$

The electron beam power P_b and the threshold resonance power P_b^{th} are

$$P_b=m_0(\gamma_0-1)c^2N_z v_{\parallel 0} \tag{46}$$

$$P_b^{Th}=-\frac{\bar\omega^2\beta^2_{\parallel 0}\gamma_0(\gamma_0-1)}{Q_L\bar L a}\left(1-\frac{m^2}{\bar k_n'^2}\right)J^2_m(\bar k_n')\frac{m_0^2c^5}{4\times 10^{10}e^2} \tag{47}$$

where

$$a=\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\sin^2\frac{\Delta}{2}\Big[h^H_{m,l}\bar{\tau}_0\beta^2_{\perp 0}\Big\{(\bar{\omega}^2-\bar{k}^2_{/\!/})\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\left[\frac{1}{2}\cos\frac{\Delta}{2}-\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)\right]-\bar{k}_{\parallel}\bar{L}^{-1}\left[\cos\frac{\Delta}{2}-\left(\frac{1}{\Delta}+\frac{1}{\Delta'}\right)\right]\Big\}+q^H_{m,l}\left(\frac{\bar{\omega}-\bar{k}_{\parallel}\beta_{\parallel 0}}{\Delta}-\frac{\bar{\omega}+\bar{k}_{\parallel}\beta_{\parallel 0}}{\Delta'}\right)-\varepsilon_H\left(\frac{1}{\Delta}-\frac{1}{\Delta'}\right)\bar{\Omega}_0\Big]$$

$\bar{L}=L/a$; $\bar{\tau}_0=\tau_0 c/a$; $\bar{\omega}=\omega a/c$; $\bar{\Omega}_0=\Omega_0 a/c$. The power unit is kW. M_0, e and c are in Gauss units.

The resonance current I_{st} and frequency offset $\Delta\omega/\omega$ are

$$I_{st}=-\frac{5\bar{\omega}^2\beta_{\parallel 0}\gamma_0}{2Q_L\bar{L}a}\left(1-\frac{m^2}{\bar{k}'^2}\right)J^2_m(\bar{k}'_m)\,\frac{m_0}{e}c^2 \tag{48}$$

$$\frac{\Delta\omega}{\omega_0}=-\frac{P_{ei}}{2\omega W},\quad \Delta\omega=\omega-\omega_0 \tag{49}$$

REFERENCES

1. M. Friedman et al., Phys. Rev. Lett., Vol. 31, pp. 752-755, 1973.
2. E. Ott et al., IEEE Trans., Vol. PS-3, pp. 1-5, 1975.
3. P. Sprangle et al., IEEE Trans., Vol. MTT-25, pp. 528-544, 1977.
4. Liu Shenggang, Scientia Sinica, No. 5, p. 905, 1979.
5. K.R. Chu et al., Phys. Fluids, Vol. 21, No. 3, pp. 461-466, 1978.
6. K.R. Chu, Phys. Fluids, Vol. 21, No. 12, pp. 2354-2364, 1978.

12553/12712
CSO: 4008/260

APPLIED SCIENCES

PHASE DISTRIBUTION OF PHASED ARRAY LENS ANTENNA

Beijing DIANZI XUEBAO [ACTA ELECTRONICA SINICA] in Chinese Vol 11 No 3, May 83 pp 73-78

[Article by Xiang Zhengping [4161 2973 1627] of Nanjing Institute of Electronic Technology]: "Study of Phase Distribution for Phased Array Lens Antenna"; paper received April 82, finalized July 82.]

[Text] Abstract: A passive lens based on the principle of prismatic refraction was introduced to amplify the scan angle of a phased array antenna. The lens might consist of an array of curved surface of revolution or a dielectric material. Formulas to calculate array lens phase distribution, configuration curve of the dielectric lens and aperture phase distribution of a planar phased antenna. When used in combination with a planar phased antenna, this lens enables a planar phased antenna beam to achieve hemispherical coverage.

I. Introduction

Since the 1960's, planar phased antennas have been rapidly developed and widely used. However, its scan angle has to be less than 70°. Otherwise, its linearity will decrease noticeably. Therefore, if hemispherical scanning is required, it is necessary to use many planar arrays. In order to expand the scan angle of a planar array, a combination of phased array and lens can satisfy this requirement[1].

The basic principle of this lens is similar to prismatic refraction in optics. As shown in Figure 1, the angle between the reflected light and the z-axis is increased relative to the angle θ when the light is reflected by the prism. $K=\Psi/\theta$ is the magnifying factor and $\Psi>\theta, K>1$. This effectively magnifies the incident angle of the light beam. The phased array-lens combination is composed of a feed planar phased array and an array of passive lenses[1]. As shown in Figure 2, 1 is a passive lens, e is a transmitting and receiving antenna element with a fixed phase distribution and F is a feed array which is an ordinary planar phased array. In this work, the passive lens may be an array of any surface of revolution. As a special case, it may be spherical. The phase delay is fixed in the lens array in order to lower cost.

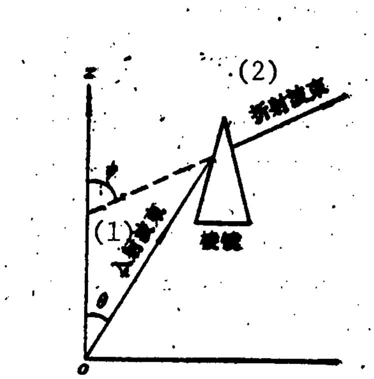


Figure 1. Principle of Prismatic Refraction

Key:
1. incident beam
2. refracted beam

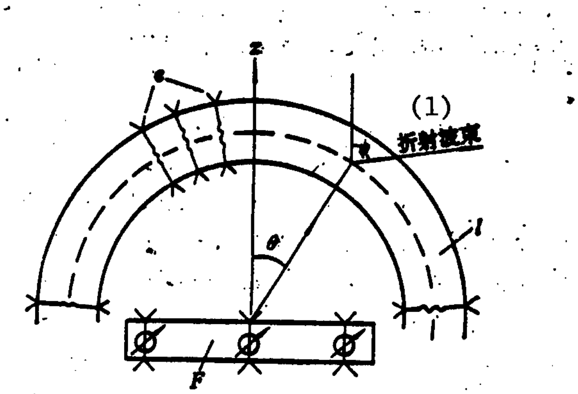


Figure 2. Schematic Diagram of Array Lens Antenna

Key:
1. refracted beam

II. Formulas for Fixed Phase Delay

When an electromagnetic wave propagates in a medium, the direction of propagation is curved toward the lagging phase. Therefore, it is necessary to provide a phase gradient by the lens in order to increase the scan angle of the planar array. The wave transmitted from the center of the feed planar array obtains an additional phase after passing through a point on the lens to avoid a change of direction of propagation. To this end, it is necessary to add a fixed phase delay on the lens element. In the following, the formulas for fixed phase delay are derived. Let us assume that the lens array is a curved surface of revolution formed by a curve $r(\theta)$ around the z-axis. Antenna elements are evenly distributed on this curved surface. We require that the lens should have a magnifying effect on the scan angle in the θ direction. The scan shall remain symmetric in the azimuth direction.

As shown in Figure 3, the wave front $\overline{QS}$ of the reflected beam is an equi-phase plane. In order to make the field at S equal to that at Q, the phase difference caused by $\overline{MS}$ should be equal to the phase increment $d\phi$ at point Q relative to M, i.e.

$$\frac{2\pi}{\lambda}dr + d\phi = \frac{2\pi}{\lambda}\sin\beta \cdot dS \qquad (1)$$

In addition

$$\beta = \psi - \theta + i \quad \text{and} \quad \psi = K\theta, \text{ therefore}$$
$$\beta = (K-1)\theta + i \qquad (2)$$

where i is the angle between the radial vector and the normal of the curved surface and β is the angle between the refracted beam and the normal direction. Furthermore, there are:

$$\operatorname{tg} i = dr/r\,d\theta \tag{3}$$

$$dS = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2} \cdot d\theta = \frac{r\,d\theta}{\cos i} \tag{4}$$

From eq. (1) and (2), we have

$$\sin\beta = \sin(K-1)\theta\cos i + \cos(K-1)\theta\sin i \tag{5}$$

By substituting eq. (5) and (4) into (1), we get

$$\frac{d\Phi}{d\theta} = \frac{2\pi}{\lambda}\left\{r\sin(K-1)\theta - [1-\cos(K-1)\theta]\cdot\frac{dr}{d\theta}\right\} \tag{6}$$

III. Formulas for Aperture Phase of Feed Planar Phased Array

In the previous section, we assumed that the electromagnetic wave was emitted from the center of the array and considered that the dimension of the feed array is much smaller than that of the lens. This is correct in the derivation of formulas for the fixed phase delay case. In reality, however, the passive lens depends on the antenna element to generate the wave. If the lens is located in a highly focused beam, then the radiated area is too small to form a secondary wave. In order to expose more area of the lens, it must be placed near the feed array. Because the index of refraction of the lens differs from point to point, it is not possible to form a beam in the k0 direction if we still use a linear phase distribution as that for the scan angle of a planar feed array beam. A non-linear phase compensation will have to be added. Let us now analyze the situation from the angle of a receiving array[2]. As shown in Figure 4, it is the sectional curve of any arbitrary surface of revolution and there is a phase delay only latitude-wise. The magnitude is determined by eq. (6).

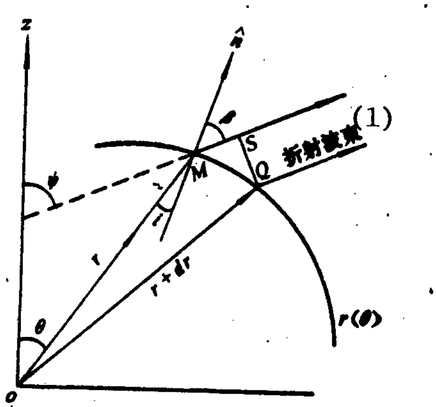


Figure 3. Geometry of Surface of Revolution Array Lens

Key:
1. refracted beam

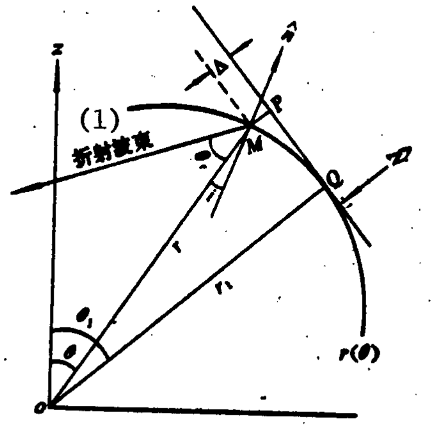


Figure 4. Coordinate to Calculate Feed Array Aperture Phase

Key:
1. refracted beam

From the path difference from the wave front to the curved surface

$$\Delta=(r-r_1)\cdot\hat{N} \tag{7}$$

where r_1 is the radial vector at the intersect between the curved surface to the wave front, r is the radial vector at any point on the curved surface and $\hat{N}$ is the unit vector in the direction of propagation, we can determine the phase difference caused by this path difference

$$\phi=\frac{2\pi}{\lambda}(r-r_1)\cdot\hat{N} \tag{8}$$

In order to determine the path difference, it is necessary to find the coordinate for the intersect (θ_1,φ_1). If the curved surface is expressed by the following equation

$$r=x(\theta,\varphi)\hat{i}+y(\theta,\varphi)\hat{j}+z(\theta,\varphi)\hat{k} \tag{9}$$

then the components of the normal vector at any point on the axes of a rectangular coordinate are:

$$n_x = \begin{vmatrix} \dfrac{\partial y}{\partial \theta} & \dfrac{\partial z}{\partial \theta} \\ \dfrac{\partial y}{\partial \varphi} & \dfrac{\partial z}{\partial \varphi} \end{vmatrix}, \quad n_y = \begin{vmatrix} \dfrac{\partial z}{\partial \theta} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial z}{\partial \varphi} & \dfrac{\partial x}{\partial \varphi} \end{vmatrix}, \quad n_z = \begin{vmatrix} \dfrac{\partial x}{\partial \theta} & \dfrac{\partial y}{\partial \theta} \\ \dfrac{\partial x}{\partial \varphi} & \dfrac{\partial y}{\partial \varphi} \end{vmatrix} \qquad (10)$$

The unit normal vector is $\hat{n} = \dfrac{n_x + n_y + n_z}{\sqrt{n_x^2 + n_y^2 + n_z^2}}$ (11)

Then, we make $\hat{n}=\hat{N}$ to obtain (θ_1, φ_1). Hence, the radial vector r_1 at the intersect could be determined to find the phase ϕ at any point on the curved surface.

According to the principle of electromagnetic wave propagation, the direction of the beam (which is the direction of phase gradient) is perpendicular to the phase plane in a homogeneous medium. Because the phase is distributed on a curved surface, the gradient is also calculated based on surface parameters. In a spherical coordinate with a surface of revolution, the components of the gradient are in the directions of $\hat{n}$, $\hat{\theta}$ and $\hat{\varphi}$. Let us make the fixed phase determined by eq. (6) be ϕ_1 and that from eq. (8) be ϕ_2, then the phase distribution function of each antenna element is $\phi_1+\phi_2$. In order to determine the direction of the beam, we attempted to find the gradient of the above phase function.

$$\left.\begin{aligned} \Gamma_\theta &= \frac{1}{2\pi} \frac{\partial}{r\partial\theta}(\phi_1+\phi_2)' \\ \Gamma_\varphi &= \frac{1}{2\pi r \sin\theta} \cdot \frac{\partial}{\partial\varphi}(\phi_1+\phi_2) \\ \Gamma_n &= \sqrt{1-\Gamma_\theta^2-\Gamma_\varphi^2} \end{aligned}\right\} \qquad (12)$$

In order to find the path length from a point on the surface to the intersect, components of the gradient in the rectangular coordinate are obtained. They are

$$\left.\begin{aligned} \alpha_x &= (\Gamma_n \cos i + \Gamma_\theta \sin i) \sin\theta \cos\varphi \\ &\quad + (\Gamma_\theta \cos i + \Gamma_n \sin i) \cos\theta \cos\varphi - \Gamma_\varphi \sin\varphi \\ \alpha_y &= (\Gamma_n \cos i + \Gamma_\theta \sin i) \sin\theta \sin\varphi \\ &\quad + (\Gamma_\theta \cos i + \Gamma_n \sin i) \cos\theta \sin\varphi + \Gamma_\varphi \cos\varphi \\ \alpha_z &= (\Gamma_n \cos i + \Gamma_\theta \sin i) \cos\theta - (\Gamma_\theta \cos i + \Gamma_n \sin i) \sin\theta \end{aligned}\right\} \qquad (13)$$

Let us assume that a beam is reflected at M whose coordinates are

$$x_M = r\sin\theta\cos\varphi, \quad y_M = r\sin\theta\sin\varphi, \quad z_M = r\cos\theta \tag{14}$$

The equation of reflected beam is

$$\frac{x - x_M}{a_x} = \frac{y - y_M}{a_y} = \frac{z - z_M}{a_z} \tag{15}$$

The intersect of this straight line with the feed planar array (assuming it coincides with the xoy plane) can be determined by letting z=0. Therefore, the coordinates for the intersect are

$$\left.\begin{aligned} x_e &= r\left[\sin\theta\cos\varphi - \frac{a_x}{a_z}\cos\theta\right] \\ y_e &= r\left[\sin\theta\sin\varphi - \frac{a_y}{a_z}\cos\theta\right] \\ z_e &= 0, \; z_M \text{ see eq. (14)} \end{aligned}\right\} \tag{16}$$

The length between the point of refraction and the intersect is

$$L = \sqrt{(x_e - x_M)^2 + (y_e - y_M)^2 + z_M^2} \tag{17}$$

The phase caused by this path is $\phi_3 = 2\pi L/\lambda$. Thus, the total phase at the aperture in the planar array from the reception point of view is

$$\Phi = \frac{2\pi}{\lambda}(L + \Delta) + \phi_1 \tag{18}$$

When the array is used as a transmitting antenna, the negative value of eq. (18) should be used in order to form a focused beam in the ψ direction.

IV. Dielectric Type Lens Antenna

The phase delay in a lens array may be achieved by changing the thickness of the dielectric lens. As shown in Figure 5, let us assume that the wave is emitted from the origin o and is deflected at an angle by the lens. The angle

between the beam and the z-axis is amplified from θ to ψ. Because the lens is assumed to be spherical, there is no refraction. By applying the law of refraction to the lens and from eq. (2), we get

$$\eta \sin i = \sin[i + (K-1)\theta] \tag{19}$$

where μ is the index of refraction of the medium. After expanding the above equation, we get the following differential equation based on eq. (3):

$$\frac{dr}{r} = \frac{\sin(K-1)\theta d\theta}{\eta - \cos(K-1)\theta} \tag{20}$$

By integrating eq. (20) with the initial condition: $\theta=0$ and $r=R$, we get the equation of the exit surface of the lens.

$$r = \frac{R}{(\eta-1)^{1/(K-1)}}[\eta - \cos(K-1)\theta]^{1/(K-1)} \qquad (K \neq 1) \tag{21}$$

Lenses of different magnifying factors may be designed by choosing a different constant K.

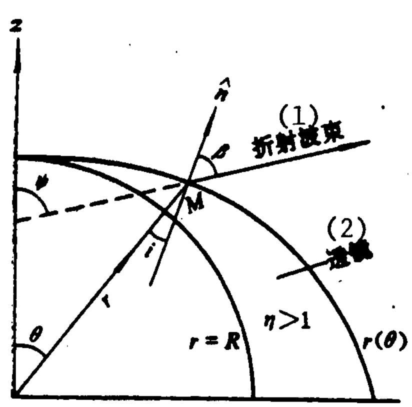


Figure 5. Dielectric Lens

Key:
1. refracted beam
2. lens

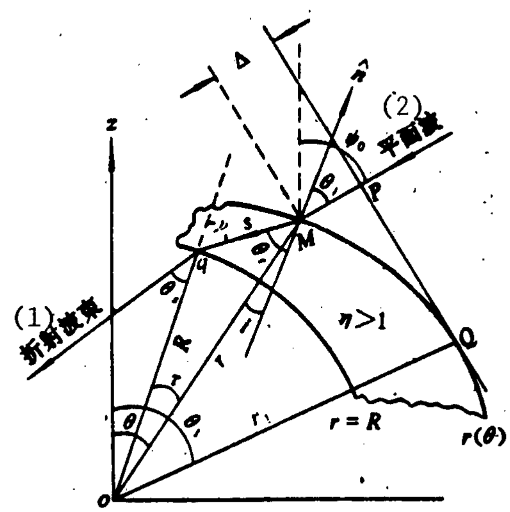


Figure 6. Schematic Diagram of Aperture Calculation of Cylindrical Dielectric Feed Array

Key:
1. refracted beam
2. plane wave

V. Calculation of Aperture Phase in Cylindrical Dielectric Lens Feed Arrays

Let us determine the angular relationship based on Figure 6. We are still looking from the viewpoint of a receiving array. Let us assume that the incoming wave is planar, it propagates at an angle ψ_0 with respect to the z-axis and i is the angle between the radial vector of the emerging surface and the normal i direction. From Figure 6 we know that the incident angle is:

$$\theta_i = \psi_0 + (i - \theta) \tag{22}$$

The angle of refraction can be obtained based on the law of refraction:

$$\theta_r = \sin^{-1}\left[\frac{1}{\eta}\sin\theta_i\right] \tag{23}$$

According to the relation of an arbitrary triangle oMq: $r/\sin(\pi-\gamma)=R/\sin(\theta-i)$ we get

$$\gamma = \sin^{-1}\left[\frac{r}{R}\sin(\theta_r - i)\right] \tag{24}$$

With respect to the incident surface of the lens, r is the incident angle. Therefore, the angle between the refracted beam and the radial vector of the inner surface of the lens is

$$\theta_r = \sin^{-1}\left[\frac{\eta r}{R}\sin(\theta_r - i)\right] \tag{25}$$

In order to determine the entire path length from the incoming wave plane to the aperture of the feed array, we must calculate three length segments, i.e. the path difference from the wave front to the exit surface of the lens, the distance s from M to q and the distance L from q to the feed array. Δ can be determined based on the following formula. First, let us find the tangential angle θ_1.

$$\Delta = r_1\cos(\psi_0 - \theta_1) - r\cos(\psi_0 - \theta) \tag{26}$$

where r_1 is the length of the radial vector at the intersect. s can be found based on the relation of the triangle oMq

$$s=\sqrt{r^2+R^2-2rR\cos\tau} \tag{27}$$

where $\tau=\gamma+i-\theta_r$. The formula for L is

$$L=R\cos(\theta-\tau)/\cos(\theta+\theta_s-\tau) \tag{28}$$

After the direction of the incident wave ψ_0 is known, the above formulas can be used to calculate the path differences. The total phase can be obtained by using a point to point method

$$\Phi=\frac{2\pi}{\lambda}(\Delta+\eta s+L) \tag{29}$$

With respect to the feed array coordinate

$$x=R\cos(\theta-\tau)\cdot \mathrm{tg}(\theta+\theta_s-\tau)-R\sin(\theta-\tau) \tag{30}$$

we are able to obtain a phase distribution in the feed array as a function of x.

With respect to a discrete curved array, the fixed phase of each antenna element in the formulas given in Sections 3 and 4 shall have the value in its corresponding coordinate.

The author wants to thank Professor Bao Jiashan [7637 1367 0810] and Associate Professor Zhang Shanjie [1728 0810 2638] for reviewing this paper.

REFERENCES

1. USP: 37J5815, 1973.

2. Xiang Zhengping, Analysis of Dome Antenna, XIANDAI LEIDA [MODERN RADAR], Vol 2, No 6, 1980.

12553/12712
CSO: 4008/260

APPLIED SCIENCES

INTRODUCTION TO REAL-TIME DISTRIBUTED COMPUTER NETWORK SOFTWARE

Beijing JISUANJI XUEBAO [CHINESE JOURNAL OF COMPUTERS] in Chinese Vol 7, No 1, Jan 84 pp 1-9

[Article by Cao Dongqi [2580 2639 0796], Institute of Computing Technology, Chinese Academy of Sciences: "Overall Design of RDC Network Software"; paper received 23 November 1982]

[Text] Abstract: RDC-net (Real-time Distributed Control network) is an experimental computer network with real-time distributed control and resource-sharing. It is built in two-level structure and distributed configuration. A survey of the scheme of RDC network software development is given in this paper. The work in the first phase of the project is also briefly described.

The Real-time Distributed Computer Network (RDC Net, for short) is an experimental computer network designed in two-level structure and distributed configuration for real-time distributed control and resource sharing. After its construction it can support many large, medium, mini, and microcomputers and through many means of communication form a network in any topological configuration for data communication and resource sharing, provide distributed processing and real-time control functions, and depending on the needs of users, be used in such fields as information transmission, data processing, process control, information retrieval, enterprise management, and military command. The development of RDC Net software was a large-scale software project. The overall development plan was called the RDCP plan and was divided into three phases, called RDCP1, RDCP2, and RDCP3. This article is devoted to the overall conception of RDCP and a general picture of RDCP1 network software. The following six articles in this issue will deal with the six network software systems developed in the first phase of the project.

I. Overall Conception of the RDC Network

1.1. Network Configuration

Depending on the physical conditions of manpower, time, and equipment, the RDC network will be constructed on the principle of from small to large, and from base levels to upper levels. Figure 1 sketches the RDCP1, RDCP2, and RDCP3 network configurations. These configurations are upwardly compatible and expandable. It is easy to expand the RDC network scale in a planned and

measured way, to improve and perfect network functions, and to expand the network's range of applications.

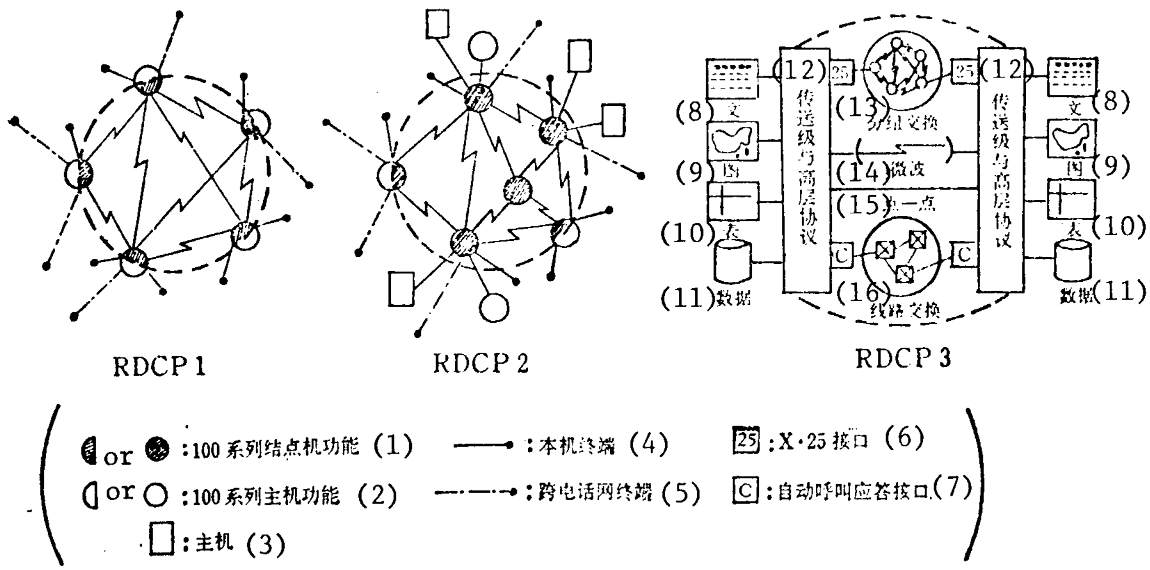


Figure 1. Diagram of Network Structure in Each Phase

Key:
1. 100 series node computer function
2. 100 series host computer function
3. Host computer
4. Local computer terminal
5. Crossover telephone network terminal
6. X·25 interface
7. Automatic call-respond interface
8. Text
9. Graphic
10. Chart
11. Data
12. Transport level and higher level protocol
13. Packet exchange
14. Microwave
15. Point-to-point
16. Line exchange

The characteristic features of the RDCP1 network configuration are: 1) Many DJS-100 series computers can be linked together in a network in any distributed configuration using communication lines; 2) in addition to assuming the functions of node computers, the 100 series computers in the network also have host computer functions; 3) the terminals in the network are simple terminals (for example, five-bit or seven-bit keyboard printer and keyboard display), they can be local terminals directly linked to the node computer and they can also be remote terminals linked to the node computer by the telephone network. The 100 series computers were selected as the node computers and host computers of RDCP1 because current domestic output is high, they are the most widely used, and the software system analysis is very clear. This choice will help us get the fastest results possible and presents a large ground for promoting applications. With the development and maturation of other domestically produced computers, when necessary, adding or shifting to more suitable computers as node computers may be considered.

The main characteristic features of the RDCP2 network configuration is the resolution of three interface problems on the basis of the RDCP1 network configuration, i.e., interfacing with large and medium host computers or high grade 100 series computers, interfacing with microcomputers and Chinese character terminals, and changing the asynchronous communications interface between node computers to a synchronous interface. In RDCP2, the 100 series computers can function independently as node computers and host computers or have two functions at the same time. When this part of the project is completed, the RDC network will form a two-level computer network of preliminary scale. That is, it will be made up of a reliable and highly efficient communications subnetwork and a well equipped and full function resource subnetwork.

The characteristics of the RDCP3 network configuration is that the problem of diversification upward and downward is resolved. The aim of downward diversification is to make the RDC network a communications subnetwork that can adopt many available communications methods, provide reliable and efficient communication for higher levels, and can adapt to the demands of different applications. For example, in addition to adopting the X.25 packet exchange, adoption of such communications methods as direct point-to-point links, line exchange, microwave communications, and communications satellites may also be considered. The aim of upward diversification is to permit many resources to join the network and support many network applications. For example, to adapt to applications demands in such areas as graphics, mapping, Chinese character processing, and list processing, RDCP3 should permit connecting the relevant equipment, formulating the relevant network protocols, and writing the relevant software system (such as text, graphics, and spreadsheet processing systems).

1.2. Network Protocol Levels

The RDC network includes two basic entities; one is the user and the other is the resources. The user communicates directly or indirectly in some dialogue form with the user interface which carries out the assigned task on behalf of the user; the resources, interfaced with a specific resource (including hardware resources, software resources, and data resources) provides the relevant service to the user. Generally speaking, the entities in the network communicate with each other and interact according to network protocols. A protocol means the arrangement which must be observed when a dialogue is carried out between specific entities and is ordinarily made up of syntax, semanteme, and exchange rules. The RDC network protocol adopts the level structure as illustrated in Figure 2. It is worked out in interconnected modules according to the ISO development system and on the basis of our specific circumstances. To be specific, they are made up of seven protocols: physical level, link level, network level, transport level, meeting level, expression level, and application level, whose functions are generally described below:

First level (physical level): The physical connection between data link entities. RDCP1 is temporarily using asynchronous communications interfaces to link with the 100 series computers. Later RDCP2 will use synchronous interfaces.

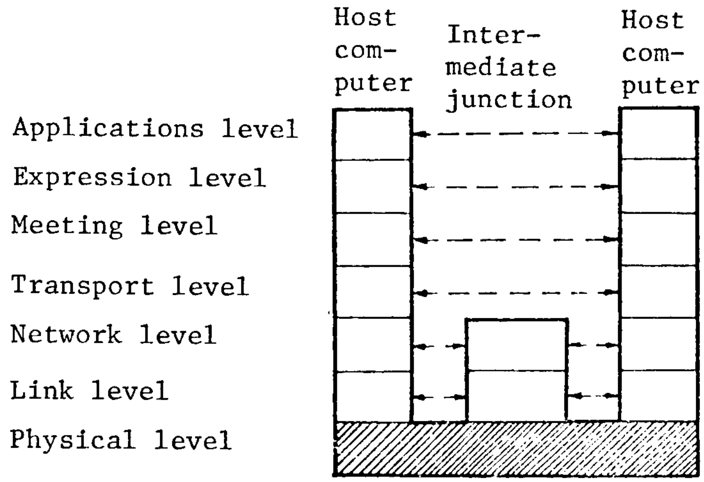


Figure 2. RDC Network Protocol Level Model

Second level (link level): Error-free data transmission between neighboring node computers. Uses CCITT X.25 non-equalized link access protocol (LAP) and the equalized link access protocol (LAPB).

Third level (network level): Provides transport level with set up, maintenance, and remove network connect functions and data transmission and control functions. The second period project uses CCITT X.25 packet level protocol; RDCP3 will consider adding line exchange and other communications methods.

Fourth level (transport level): Achieves end-to-end connections independently of basic network communications, provides process-to-process communications for different processes on the computer systems in the network, and maintains clear data transmission between them. It uses the internetwork end-to-end transport protocol proposed by ISO.

Fifth level (meeting level): Supports interaction between cooperating application entities. Control functions provided include control of set up, discrimination, and removing of meeting relations between application entities and synchronous control and boundaries of data exchange, to help ensure that safe, reliable, and orderly dialogue can be conducted between application entities. There are three types of meeting form: unidirectional, staggered bidirectional, and simultaneous bidirectional.

Sixth level (expression level): This level provides an overall formatted service to help applications level entities analyze and interpret the content and meaning of the data exchanged. The typical expression level protocols include dumb terminal protocol, file transmission protocol, remote job entry protocol, and electronic mail protocol. They are responsible for the coordinated management of various operations which include recording, exchange, storage, search, display and control of various data structures.

Seventh level (applications level): This level consists of the various types of applications systems and the user-written programs. They can call up the

various network service functions through software at the expression level. As far as the range of applications is concerned, this level has a wide range and number of protocols used for supporting network resource sharing, typical task activity, specific enterprise management, and safe operation of network applications systems.

Of these seven levels, levels 1-3 constitute the network communications service level, levels 5-7 constitute the network applications service level and the fourth and transport level links these two groups of levels together. One of its important functions is to choose the optimum communications service method for carrying out relevant network activity on the basis of the service demands and parameters of the network applications service level.

1.3. Principles of Writing Network Software

In the RDC network software development process, we decided on principles of simplicity, practicality, reliability and effectiveness and on this basis tried to make the system expandable and to make it possible to delete elements so that it would be more adaptable. For this reason, in the software system structure we adopted design method by levels, modules, and processes, and planned to produce a system which would make the network software easy to edit and expand to adapt to the function demands and differences in resources and equipment of the various stations and to the developments and changes of each phase of the project. The overall functions of the RDC network are implemented through various network protocol software and network applications systems. Network protocol software was developed by levels from bottom to top taking the seven network protocol level models as natural levels. Between all the levels there is a partially ordered unidirectional dependency relationship in which each level is related only to neighboring levels: the lower level is the foundation of the upper level and provides services to the upper level; the upper level is implemented through functions supplied by the lower level and for the higher level provides functions which are more powerful. There is a clear interface relationship stipulated between level and level and when a level is expanded or revised, as long as the interface relationship is not changed, the functions and structure of the other levels are not affected. Generally speaking, each level is made up of a module with unitary functions, clear interface, and moderate size and they cooperate among themselves through a unified form of transfer and communications. Each module in a level usually can also be further subdivided into a certain number of levels for editing in accordance with the transfer relationship. One important characteristic of network activity is the high degree of coincidence. To ensure this coincidence, the network protocol software or network applications system of each level in the RDC network is implemented through one or many processes.

II. General Picture of RDCP1 Network Software

RDCP1 network software was developed on the DJS-100 series computers and includes link level software, packet level software, transport level software, packet assembly/disassembly software, mini relational data base system and remote job entry system. The first four systems were designed in accordance with the four lowest protocol levels in protocol level model or relevant

international standards. Taken together they make up a network communications service level which has stable functions. The last two systems are the two network applications systems. They operate under the support of the first four systems. Since there are no functions which independently provide higher level network protocols in the first phase of the project, the latter two systems have some simple and necessary meeting level and expression level functions. Figure 3 gives the positions occupied by the six systems in the network level model.

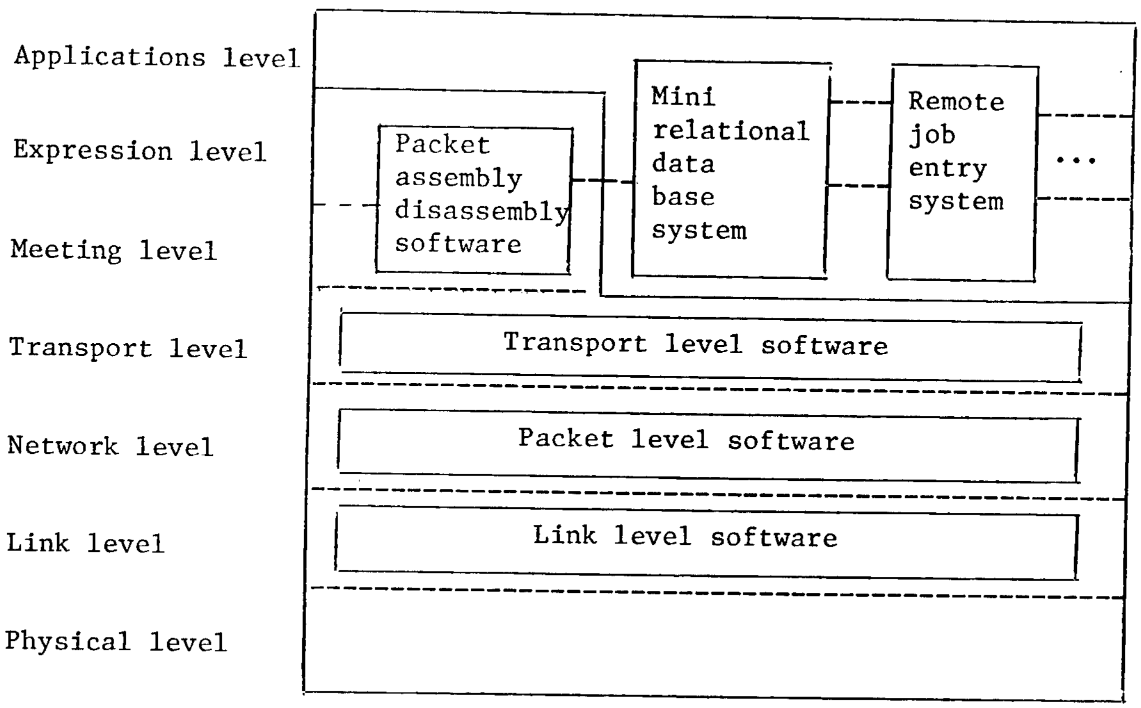


Figure 3. RDCP1 Network Software by Levels

So that the user will have the vigorous support of the RDOS system when developing new network applications, we decided on a policy of developing six network software systems under the support of the RDOS system and making them subsystems under the control of the RDOS system. However, to adapt to the network environment, we also made some necessary revisions and expansions in the RDOS system. Below we will first introduce the revisions and expansion of the RDOS system, then reintroduce separately the four network protocol software and two network applications systems.

2.1. RDOS System Revisions and Expansions

As stated above, RDCP1 network software was developed under the support of the RDOS system. Specifically, utilizing the multi-task management functions of the RDOS system to support communications, and synchronous and coincidental operations between many processes; carrying out the necessary file operations through the various functions provided by the system used by the call command

in the RDOS system; executing exchange and link commands to cut between systems within the network software; establishing timed or time-delay tasks to implement time control and carry out cyclical work. However, to adapt to the work characteristics of the network environment, so that the network software manages communications control and data transmission by itself, to implement various types of communications control rules, and to lighten the work volume of the upper level network protocols and the network applications system, we made the following revisions and expansions to the RDOS system.

(1) Revised the QTY drive program to add a call-respond module.

To support link level protocol between DJS-100 series computers and to support communications between the terminals by the network software and computers, and for basic transmission control rules between the computer and long-distance telephone network terminals, we changed the QTY asynchronous multiplex communications controller into a corresponding user defined device and self-editing interrupt processing program. In addition, we also added a call-respond module for timed task operations to monitor and manage the telephone line call bell signal and carrier signal of the communications process.

(2) Added a queue management function.

To facilitate using a method of registering processing units between processes for requests back and forth and to implement mutual communication and cooperation, we added a processing unit queue for each process in the network software and provided the three queue primitives PUT, GET, and GETS (see Table 1).

(3) Added buffer management functions.

To support the user program and network software transmitted data and implement network protocol, buffer management functions were added to the foundation of the RDOS system. Buffers are divided into large and small: each 40-digit long large buffer is used for transmitted data; each 10-digit long small buffer is used for transmission protocol commands. Correspondingly, the primitives AG, AL, and RL (see Table 1) were provided so that the user could flexibly use and manage the relevant buffer.

(4) Exclusive operational primitives.

To implement mutual exclusion between processes, we introduced N access values as an 0 or 1 signal volume, written as S1, S2,..., SN and provided the two operational primitives PS and VS. Actually, the value of the signal volume reflects the user situation of the resources: when the value is 1, it means they are already occupied; when the value is 0 it means that they are not yet occupied. Thus, in terms of processes competing for resources, the mutual exclusion function will be achieved if a PS operation is carried out before entering the stage of mutual exclusion and a VS operation is carried out on withdrawing from the stage of mutual exclusion.

Table 1. Network Software Expansion Primitives and Their Functions

Class	Primitive	Functions
Queue management	PUT(P,M)	Combines the number of the process called by the primitive and the information content M, registers it in the process queue of the designated process P. If the queue is empty when the process is registered, then it will activate the corresponding receiving process P.
	GET	Calling the process of this primitive the first process unit is taken from its own processing unit queue so as to carry out the corresponding process in the service request issued. If the process unit queue is empty an error is returned and the caller processes it by itself.
	GETS	Similar to the GET primitive but differs in that if the process unit queue is empty, it does not return an error but registers the process called by the primitive until another process is sent by a PUT primitive at which time it is reactivated and continues work.
Buffer management	AG(A,S)	Requests a large buffer. If there is a buffer that can be provided, after the return in status S indicates that there is a buffer and in A gives the first address in the buffer assigned. If there is no buffer, it is so indicated in status S.
	AL(A,S)	Functions similar to AG but requests the assignment of a small buffer.
	RL(A)	When the buffer is no longer needed, this command can be called to release the buffer at the address indicated by A.
Dismantling operations	PS(SEM)	Determines whether the value of the corresponding signal SEM is 0: if the value is 0 then the signal is set to 1, then it returns and continues execution; if the value is 1, then the process called by this primitive is registered in the corresponding signal queue.
	VS(SEM)	Sets corresponding signal to 0 and if the signal's queue is not empty, then activates the process to which it is registered.

P = expression process signal; M = expression information content; A = expression buffer address; S = expression buffer distribution; SEM = expression signal number.

2.2. Network Protocol Software

RDCP1 network protocol software is made up of link level software, packet level software, transport level software and packet assembly/disassembly software. In view of the fact that we hope to be able to operate with the support of low grade 100 series computers and a simple operating system in order to give them a broader range of application and support multiple network applications systems and user programs, when we wrote this protocol software, we used assembly commands as much as possible and used few high level operation system commands. Below is a brief introduction to these protocol software functions and their makeup.

(1) Link Level Software (LL)

This software was implemented in accordance with the CCITT X.25 link level protocol and the basic rules of ISO with revisions made on the basis of specific situations. It is situated above the physical level and is the basic protocol software of the lowest level in the network level model. It is primarily composed of the interrupt module, computer link management module, and the terminal link management module. The interrupt module is responsible for managing the information transmitted on the communication lines and its control. The computer link management module supports error-free transmission between neighboring computers and carries out HL, a link level protocol similar to the CCITT X.25's. And, the terminal link management module supports communication between terminals and computers and is responsible for carrying out BL, the basic transmission control rules between remote terminals and the computers.

(2) Packet Level Software (PL)

This software carries out CCITT X.25 packet protocols: it reuses the data links provided by the link level and provides virtual circuit service between two DTE's. In the process of setting up virtual circuits, the packet level software uses the fixed routing method to select the network path. After a virtual circuit has been set up, it is responsible for managing the transmission of data packets and interrupt packets, executing flow control, reset, and restart functions up until the data transmission stage is concluded and the virtual circuit is disconnected.

(3) Transport Level Software (TS)

The transport level software, in line with the internetwork end-to-end protocols proposed in IFIP WG 6.1, selected this subset to implement RDC net transport level protocols with the aim of providing communications functions between processes on the net. This software establishes many xinkou [0207 0605], manages correspondences between them and the user processes, and on the foundation of the virtual circuits provided at the packet level, sets up, maintains or closes the links between xinkou, depending on user requests, transmits data or telegrams over the link, and can cut to another xinkou on an already established link.

(4) Packet Assembly/Disassembly Software (PAD)

This software carries out the functions stipulated in CCITT X.3 and X.28 proposals with the aim of permitting simple terminals (such as keyboard printer terminals and keyboard display terminals) to enter the net and to implement the mutual transfer between terminal input/output character flow and transmission packet flow within the net. Based on instructions issued by the end user it can set up, reset, or remove connections between the terminal and the corresponding service program, transmit data information or interrupt information, and display or revise terminal specific data (such as whether or not a filler character should be added after the longest line or a CR).

Figure 4 illustrates the data flowchart and command flowchart between the above-described network protocol software. In the data flow, when the data information issued by the user process or terminal operator goes through network protocol software from top to bottom levels, it should carry out in sequence transformation of character flow, message, interconnected packet sequence, independent packet sequence, and frame sequence; on the other hand, when information crosses over the communications network and goes through the destination network protocol software from bottom level to top level, it goes through the corresponding reverse conversion in sequence and is finally delivered to the user process or end user at the other end. In the control flow, the network service request required by the user process or end user is carried out by the conversion of the level commands PAD, TS, TS-PL, PL-LL and I/O or their reverse conversion.

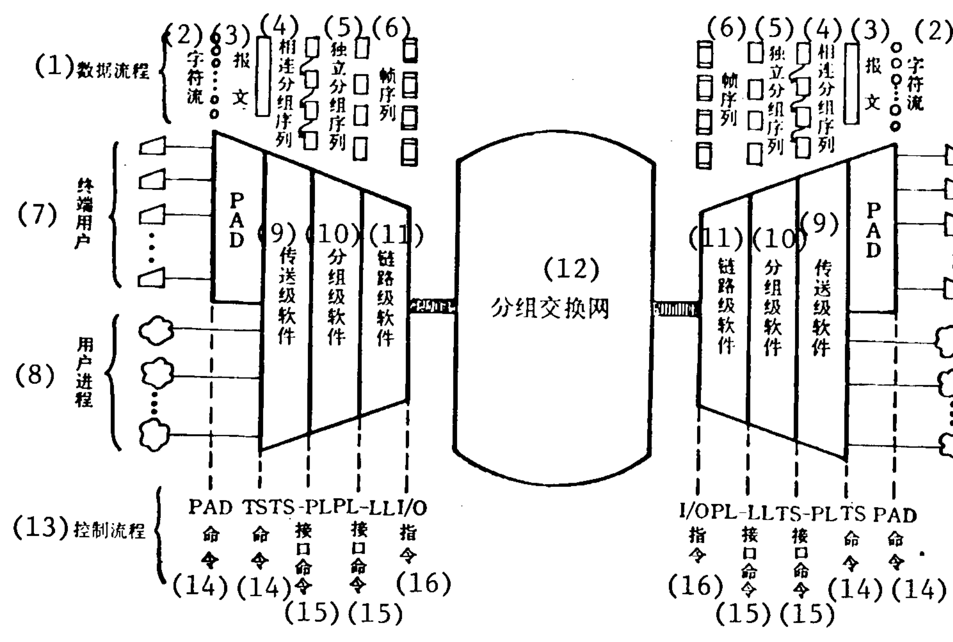


Figure 4. Data Flow and Control Flow Between Network Protocol Software

[Key on following page]

Key:

1. Data flow
2. Character flow
3. Message
4. Linked packet sequence
5. Independent packet sequence
6. Frame sequence
7. End user
8. User process
9. Transport level software
10. Packet level software
11. Link level software
12. Packet exchange net
13. Control flow
14. Command
15. Interface command
16. Instruction

2.3. Network Applications System

After the four network protocol software described in 2.2 above were set up, we created a network communications service level with stable functions and flexible application. Under its support, the user or system designer can conveniently design or write user programs which involve network activity and develop a variety of network applications systems. This is because, in the RDCP1 network software, in addition to maintaining the powerful support of the original RDOS system for user programs, the two new powerful supports of the transmitting station command and the PAD command were added for the user (see Figure 5). In short, when the user program requires network activity, through the transmitting station command it can conveniently call any program on any computer in the network and communicate and cooperate with it and jointly complete the necessary task (see Figure 5, item 1); when the user program provides network service for the end user, the end users can use the PAD command through various terminal equipment to simultaneously request the network service provided (see Figure 5, item 2); when the user program requires management resources, development work or to call on relevant files, it can use the powerful support of the RDOS system directly (see Figure 5, item 3). As an example of the actual applications value, the RDCP1 developed two network applications systems: a mini relational database and remote job entry system. They are briefly described below.

(1) Mini Relational Database System (MRDS)

This is a mini relational database system under the network environment. It consists of a network on-line search subsystem and a data base manager subsystem. The former is designed for the network user and can simultaneously support many user searches of relevant content of the database. The search command is commonly a combination of select, project, and join operations. The latter is designed for the data base manager for ordinary maintenance and management of the database. It is a single-user system for a non-network environment, and can be used for setting up, loading, revising, and deleting relational patterns and their values in the database. With the support of the network communications service level, if a large database is suitably divided into so many mini databases and rationally distributed in the network, then a distributed database function based on the small base as the distributed unit can be implemented.

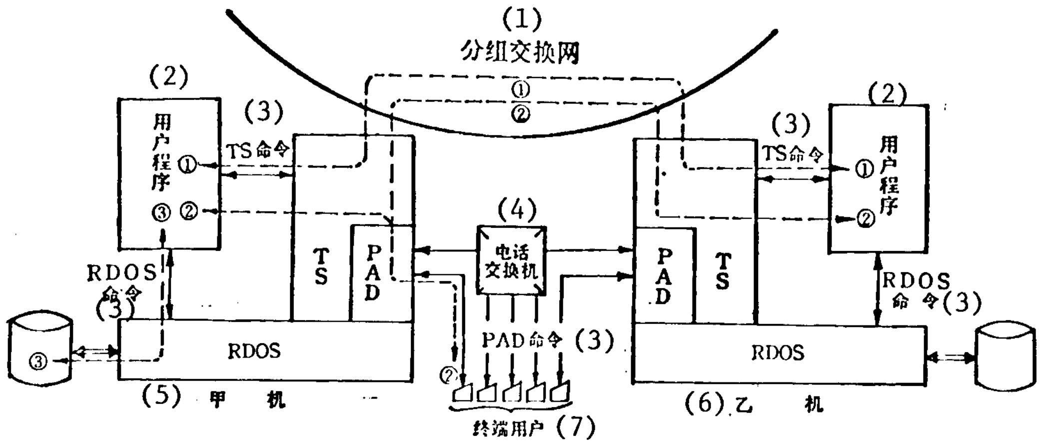


Figure 5. Support Functions of RDCP1 Network Protocol Software

Key:

1. Packet exchange net
2. User program
3. Command
4. Telephone exchange
5. Computer A
6. Computer B
7. Terminal operator

(2) Remote Job Entry System (RJE)

This system consists of account opening subsystem, operations monitoring and control subsystem, job entry subsystem, and batch processing system. The account opening subsystem is designed for the system managers and is used for opening accounts, closing accounts, statistics, collecting fees, and creating passwords. The operations monitoring and control subsystem was designed for the system operators and is used in monitoring the system operations situation and controlling system startup and stop and operations of the cutting job entry subsystem or batch processing system. The job entry subsystem is a multiuser system under a network environment and can support many network users simultaneously submitting their own jobs or requesting the results of a previous job entry from different stations in the network. The batch processing system is an original subsystem of the RDOS system but it is included under the control of the operations monitoring and control system to execute the job to be entered during the entry phase or save the operational results of the job. This phase is called the execution phase and is done under a non-network environment.

III. Conclusion

This paper has introduced the general concepts of the RDC network software development and a general picture of RDCP1 network software. RDCP development plans are made up of gradual revisions, expansions and improvements. The present RDCP1 has been completed, RDCP2 is just beginning, and RDCP3 is still only in outline. The network software developed in the first period will be further revised and perfected in actual application and we will strive to make it more practical and improve the efficiency of operation.

Some aspects of the conceptions of the second and third periods also will undergo necessary adjustments, revision, improvements and expansions in the process of being used and the key technologies also await further discussion and resolution.

RDCP1 network software was jointly developed by the comrades of the "Network Software and Database" group of the Institute of Computing Technology of the Chinese Academy of Sciences and over 10 graduate students who were pursuing studies in the group at the time. In the development process, Comrade Ma Yinglin [7456 1758 3829] provided many beneficial ideas for the overall plan and proposed the overall conception of the functions, structure and implementation of the remote job entry system; Comrades Xu Caijie [1776 2088 2638], Gao Jianyu [7559 0494 1342], Xu Zhuanyou [1776 0278 0645], and Xu Xiaqing [6079 7209 0615] and Comrade Cao Zhijiang [2580 0037 3068] did a great deal of the hardware transformation work and communications interface work for the development of the RDCP1 network software; many comrades of Digital Control Lab 1, Digital Control Lab 4, and NOVA3 Computer Lab provided us with vigorous support and enthusiastic assistance and also provided us with an excellent working environment. On behalf of the authors of the following six articles, the author would like to take this opportunity to express heartfelt thanks to all these comrades and the comrades who gave engineering support to this project.

REFERENCES

1. ISO, Reference model for open systems interconnection, International Organization for Standardization, DP 7498.

2. ISO 1745-75, ISO 3309-79, ISO 4335-79, International Organization for Standardization.

3. CCITT Provisional recommendations, X3, X25, X28, and X29 On packet-switched data transmission services, International Telecommunications Union, Geneva (1978).

4. V. Cerf, A. Mckenzie, R.A. Scantleburg, and H. Zimmerman, Proposal for an internetwork end-to-end transport protocol, IFIP, WG 6.1 (1978).

5. Cao Dongqi [2580 2639 0796], Liu Fuzi [0491 4395 3320], JISUANJI WANGLUO RUANJIAN JICHU [COMPUTER NETWORK SOFTWARE FOUNDATION], People's Posts and Telecommunications Publishing House, 1982, 7.

6. UNINETT Reports, No 1-5, Oslo.

7. D.W. Davies, D.L.A. Barber, W.L. Price, and C.M. Solomonides, Computer networks and their protocols, John Wiley and Sons, 1979.

8226/6091
CSO: 4008/243

APPLIED SCIENCES

LINK LEVEL PROTOCOL, ITS IMPLEMENTATION

Beijing JISUANJI XUEBAO [CHINESE JOURNAL OF COMPUTERS] in Chinese Vol 7, No 1, Jan 84 pp 10-19

[Article by Wu Zhimei [0702 0037 5019], Song Huanggang [1345 3562 0474], Sun Simin [1327 0934 2404], and Xu Guirong [1776 2710 2837], Institute of Computing Technology, Chinese Academy of Sciences; paper received 23 November 1982]

[Text] Abstract: The BL and HL communication protocols adopted in the link-level of RDC network are presented. Also included are the implementation of the link-level and its interface to other levels.

I. Overview

The RDC Net is an experimental computer network based on the Chinese-made DJS-100 series computer. It is divided into five levels and the link level is the basic software of the lowest level of the network. Its aim is to provide error-free data transmission between communication entities which are scattered geographically.

Currently, the RDC Net link level administers four kinds of circuits (see Figure 1). The first kind of circuit is used for linking local five-bit telecommunications equipment, the second kind of circuit is used for linking local seven-bit telecommunications equipment, the third kind of circuit is used for linking remote terminals over the Changsha City [changshi [7022 1579]] telephone network, the fourth kind of circuit is used to link the node computers in the network. The distances spanned by the first and second kind of circuits are rather short and transmissions between terminal equipment and the computer can be regarded as error free. The third kind of circuit goes over the changshi telephone network, and in the transmission process, the information is subjected to a great deal of interference, thus to eliminate this interference and ensure reliable information transmission, a basic level data communication protocol (BL) has been implemented on this type of circuit. The distances covered by the fourth kind of circuit are large and demand that simultaneous, bidirectional, and encoded data transmission can be carried out between computer and computer thus a high level data communications protocol (HL) has been implemented on this kind of circuit.

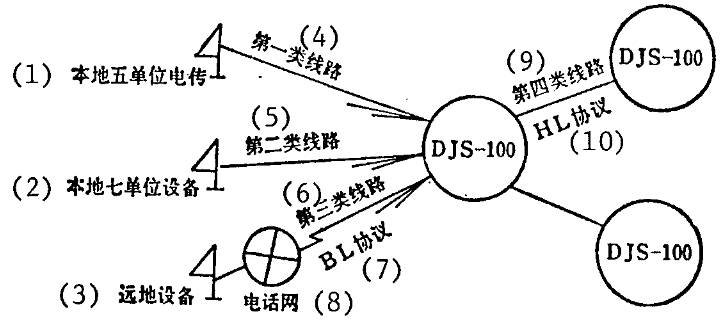


Figure 1. Connections of Various Circuits in RDC Net

Key:

1. Local five-bit telecommunication
2. Local seven-bit telecommunication
3. Remote equipment
4. First type of circuit
5. Second type of circuit
6. Third type of circuit
7. BL protocol
8. Telephone network
9. Fourth type of circuit
10. HL protocol

II. Brief Introduction to the BL and HL Data Communications Protocols

1. BL Protocol

The BL protocol adheres to the stipulations of ISO document 1745 and CCITT proposal V.24. Its basic features are as follows:

(1) The circuit connection method is point-to-point so the communications circuits can be dedicated circuits, leased circuits or public telephone exchange circuits. A modem is added at both ends of the public telephone exchange circuits or long-distance circuits.

(2) The method of signal timing is local timing, also called asynchronous, i.e., the length of each code element in a character is identical, but the time interval between characters can be of any length.

(3) The circuit work method is a full duplex circuit. At half-duplex voice transmission control transmission speed can vary between 200 and 9600 bit/sec.

(4) The character set is the international number 5 code stipulated by the ISO and the CCITT. Except for 10 transmission control characters, the other characters can be user information transmitted in the system.

(5) Error control is by cross parity check (see Figure 2). User information is expressed in a 7-bit code. Before transmission, one bit is added to each character as a parity check bit, and a block check character (BCC) is added after each group of information. At the receiving end, a parity check is done for each character and a horizontal parity check is done on each group of information. If no errors are discovered the blocks are considered correct; if an error is discovered then the transmitting terminal must retransmit that group of information.

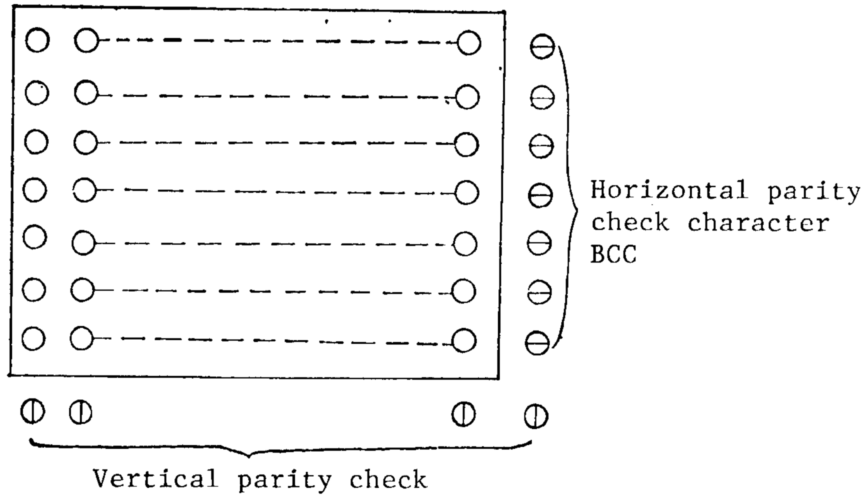


Figure 2. Cross Parity Check

(6) Flow control uses waiting type flow control; only after the first group of information has been received correctly can the next group of information be transmitted.

(7) The text format of information transmitted on the circuits is divided into two classes by function as information text or supervisory text. Information text is used for transmitting user data; supervisory text is used for transmitting control information and this control information is used for managing the process of data information transmission.

Information text format is as follows:

a	STX	STX	text	ETB	BCC
b	STX	STX	text	ETX	BCC
c	STX	STX	text	ENQ	FNQ

All of a user's data can be made up of a b-type text and can also be made up of some a-type text followed by a b-type text. The c-type text is invalid text and can appear in any position in user data flow.

Supervisory text is of the following types:

d	n ENQ	selected station (SEL) (n = source station mark)	e	ENQ ENQ	enquiry station (ENQ)
f	EOT EOT	handing over right to transmit (αα)	g	DLE EOT	disconnect (DIS)
h	DLE 0	acknowledge (ACKO)	i	DLE 1	acknowledge (ACK1)
j	DLE NAK	negative acknowledge (NAK)	k	DLE 5/12	pause (PSE)
l	DLE 3/10	request right to transmit (ββ)	m	DLE 4/14	group interrupt (BER)

From d to g are forward supervisory text, and g to m are reverse supervisory text.

(8) Communications stages and their connections. A complete data transmission process between a pair of users can be divided into five stages:

a) Establishing connection on the public telephone network. The user uses an ordinary telephone for dialup and the telephone bureau handles switching and identification.

b) Defining the data link. After the connection has been made on the public telephone network, if the unit has a switchboard, an additional switching operation has to be made to complete the circuit continuity process. Then, the party making the call becomes the primary station and issues the select sequence SEL; the party called becomes the secondary station and uses the ACKO to indicate willingness to respond, then the data transmission sending-receiving relationship is defined.

c) Data transmission. In this stage, the primary station transmits data, the secondary station gives the response; the functions of the primary station and secondary station do not change. The primary station transmits information text and if it is checked and correct, the secondary station alternately uses ACKO and ACK1 to make affirmative response. If there are errors in the information text, the secondary station uses NAK for negative response; if the secondary station is temporarily unable to receive information text, it uses PSE to request that data transmission be stopped temporarily. When the primary station receives alternate affirmative responses it transports the next information text; when the primary station receives a negative response it transmits the current information text; when the primary

station receives a temporary stop response, it transmits an enquiry test; when the primary station does not receive a response after a certain time it also transmits an enquiry test; when the primary station continues to receive a negative response N times in a row or no response N times continuously after the time limit, it should be reported to the operator that there is a problem on the circuit.

d) Concluding data link. After data transmission has ended, the primary station sends an αα text, giving up its primary station status and transfers the right to transmit to the other party. If the new primary station has a text it wants to transmit, then it returns to the third stage and carries on data transmission. If the station receiving the right to transmit does not wish to continue communication, it can transmit a DIS text cutting off the data link, then entering the fifth stage.

e) Breaking off the connection on the public telephone network. This stage is done by the telephone bureau. The process is identical to that of the ordinary telephone user breaking a connection.

2. HL Protocol

The HL protocol was designed in accordance with ISO document 3309 and ISO document 4335 and on the basis of the hardware support obtainable at the time. Its basic characteristics are as follows:

(1) Circuit connection method. Uses the point-to-point method. The communications circuits used may be dedicated circuits or leased circuits and for long-distance circuits modems must be used at both ends.

(2) Signal timing method. Partial timing method is used the same as with the BL protocol.

(3) Circuit work method. Uses four line full duplex circuit, carries out simultaneous duplex data transmission, transmission rate can vary between 200 and 9600 bit/sec.

(4) Character set. A bit string of any composition can be under the control of the HL protocol, from one node to another, but the length of the bit string must be an integer of multiples of 8.

(5) Error control method. Uses horizontal and transverse parity checking to check for errors in the transmission process (see Figure 3). Each frame transmitted includes a 16 bit frame check sequence of which the first 8 bits are horizontal check codes of information in the frame and the last 8 bits are transverse check codes of the information in the frame. At the receiving end, horizontal and transverse parity checks are done of the frame received and if there are no errors the frame is considered correct; if errors are discovered, the frame is discarded.

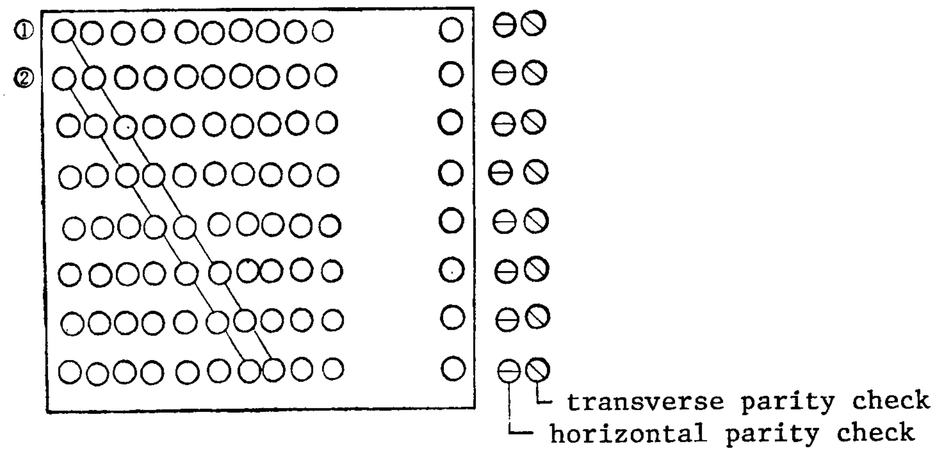


Figure 3. Horizontal and Transfer Parity Check

(6) Flow control method. Between two node computers there are simultaneously two data flows and two control flows, the control flow in one direction and the control flow of the other data flow control occupy one physical channel. Each data flow is transmitted only after the previous data frame has been correctly received.

(7) Frame format. Information transmitted on the circuit uses a uniform letter format whether it is an information frame transmitting user data or a supervisory frame transmitting control information it uses the following format:

F	L	C	I	FCS	F

F is the flag sequence (0 11 11 11 0) and takes up one byte. L indicates the frame length in bytes and takes up one byte. C indicates the type of frame and takes up one byte. FCS is the frame check sequence and takes up two bytes. The information field I only appears in the information frame and can be less than or equal to 64 bytes in length.

(8) The communications process. After the system has been loaded, the physical links are in the on state. When a user in the upper level of the network pushes data which is to be transmitted over a circuit into the link level, the link level first of all sets up a bidirectional, asynchronous, balanced link, then shifts into the data transmission stage. In the data transmission stage, each station can independently transmit information frames in sequence and each time a frame is sent, the frame sequence is increased by 1. The receiving station ignores frames checked and with errors. If the sequence number in an information frame is the frame sequence number it wants to receive, it sends back a response frame that it is ready to receive. If the sequence number in the data frame is the sequence number of the previous frame, it also sends back a response frame that it is ready to receive, but discards the information frame received; if the sequence number in the information frame is another value, it sends back a refusal response frame;

if it is unable to receive the information frame, it sends back a response frame that it is not yet ready to receive. When the sender discovers that a certain time limit has been exceeded without a response, it retransmits the current information frame. When it has gone beyond the time limit N times in a row or when there have been N negative responses in a row, the operator is notified that there is trouble on the link. If the upper level of the network temporarily does not have sufficient data for the link to transmit, the link state is still maintained in the data transmission state, and only when the upper level on the network indicates that it wants to break off the data link does the link level disconnect the data link in question.

III. Implementation of the Link Level

The link level includes five processes, one interrupt module, two communications primitives and a group of circuit characteristics tables (see Figure 4).

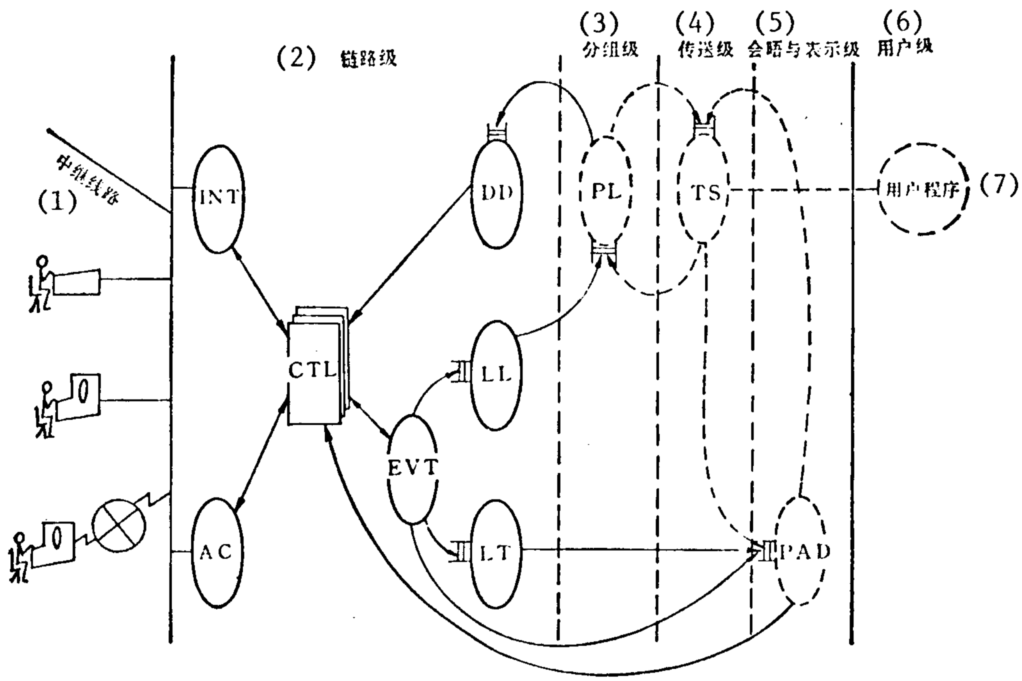


Figure 4. Make Up of Link Level

Key:

1. Interrupt circuit
2. Link level
3. Packet level
4. Transport level
5. Meeting and expression levels
6. User level
7. User program

1. Circuit Characteristics Tables CTL

A circuit characteristics table CTL is set up for each external line (including links between the DJS-100 and a terminal and links between DJS-100's) in the link level. It records such information as circuit state, data queues on

the circuit waiting to transmit, and data fields currently being received and transmitted. CTL is the basis of interrupt program work and is also the locus of the interrupt module and other local communications.

2. Communications Primitives

Communications between local area and local area relies primarily on the communications primitives PUTQ and GETQ for implementation. The function of PUTQ is to register a processing unit in the indicated queue, the function of GETQ is to access a processing unit from the indicated queue. A processing unit commonly includes the following parameters: information source, information category, data block address, and data block length. The work of each process is independent, and each processes its own work. After a task is completed, GETQ is used to get work from its own queue. When it is necessary to pass information to another process PUTQ is used to pass the processing unit to another process. This is the way these independent work processes communicate with one another and cooperate in completing larger tasks.

3. Interrupt Block INT

The interrupt block is at the lowest level in the link level. It manages input and output of character information on the four types of circuits. The hardware interrupt signal activates the interrupt program. If it is an input interrupt, depending on the indicators in the circuit characteristics table, the characters received are placed in a buffer; if it is a transmit interrupt, depending on the indicators in the circuit characteristics table, the followup characters are sent to the circuit, each time an entire information field is received or sent, a mark is made in the characteristics table in the interrupt program and information transactions of other processes and the interrupt module use the entire information field as the basic unit.

For circuits between machines, this information field is a frame; for remote terminal circuits of Changsha City [?] telephone network, the information field is a text; for circuits connecting local five-bit and seven-bit devices, the information field accepts the carriage return as the concluding character or the character string that fills the buffer. In the international number two code used by 5-bit telecommunications devices, when outputting, the interrupt program converts the ASCII character into 1-4 five-bit characters and telecommunicates them in sequence; when inputting, the interrupt program converts the 1-4 five-bit characters into the corresponding ASCII characters. After adding I through the interrupt program, the other programs no longer sense any difference between the five-bit devices and the seven-bit devices.

4. Event Analysis Process EVT

The event analysis process EVT is a timing task. It reads the circuit characteristics table and analyzes the events appearing on the circuit. If a buffer's information is transmitted and it will not be used again, the buffer is released to the free buffer pool; if a complete information field is received, then a new buffer is requested for the circuit to receive the characters. Information fields received by a local terminal are directly

transferred to the PAD process. Information text, control text, and no response beyond the time limit messages are passed to the appropriate processing unit and registered in the terminal link management processing LT queue. Frames and no response beyond the time limit messages received on circuits between machines are passed to the appropriate processing unit and registered in the intercomputer link management process LL queue.

5. Terminal Link Management Process LT

The terminal link management process LT implements a finite automaton (see Table 1). This automaton expresses the primary content of the BL protocol.

Table 1. Automaton LT

Primary station

State / Event	SDA	RPS
ACK0	LT1	LT7
ACK1	LT1	LT7
NAK	LT2	LT8
PSE	LT3	LT3
$\beta\beta$	LT1	LT7
BER	LT4	LT4
TMO	LT5	LT9
NNO	LT6	LT9
NNA	LT6	LT9
DIS	LT4	LT4

Secondary station

State / Event	RDA	SPS
ETB	LT10	LT16
ETX	LT10	LT16
ABT	LT11	LT16
ENQ	LT12	LT16
$\alpha\alpha$	LT13	LT13
NBF	LT14	LT15
YBF	LT15	LT17
DIS	LT4	LT4

Initial state

State / Event	DIS
SEL	LT18
DIS	LT4

The rows in the table indicate the possible states of circuit ports: sending data (SDA), receiving a pause request (RPS), receiving data (RDA), transmitting pause request (SPS), and initial link state (DIS).

The columns in the table indicate the events that can appear: acknowledge (ACK0, ACK1), negative acknowledge (NAK), pause request (PSE), requesting a reverse direction ($\beta\beta$), group interrupt (BER), beyond the time limit (TMO), over the time limit N times in a row (NNO), negative response N times in a row (NNA), disconnect circuit (DIS), information text (ETB, ETX), delete text (ABT), inquiry (ENQ), reverse direction ($\alpha\alpha$), have buffer (YBF), no buffer (NBF), and select text (SEL).

LTi in the table indicates the action to be taken when that event occurs in the corresponding state.

LT1: Release buffer occupied by information text currently transmitted, transmit next information text. Clear beyond the time limit counter and

no response counter, when information is transmitted transmit reverse direction text $\alpha\alpha$, change state to RDA.

LT2: Retransmit current information text, increment NAK count by 1.

LT3: Change state to RPS and transmit inquiry text.

LT4: Change state to DIS and transmit DIS to disconnect circuit.

LT5: Retransmit current information text, increment no response count by 1.

LT6: Inform operator that no response has appeared N times in a row or no response has appeared N times in a row.

LT7: Change state to SDA, transmit next information text.

LT8: Change state to SDA, retransmit current information text.

LT9: Reactivate timer, and transmit inqiry text.

LT10: Audit check sum, determine whether to send text to acknowledge positively or negatively.

LT11: Negative acknowledgement text is sent.

LT12: Transmit current ACK or NAK text.

LT13: Change state to SDA, transmit information text via this line's output queue.

LT14: Transmit request for pause, SPS.

LT15: No action.

LT16: Retransmit PSE text.

LT17: Change state to RDA.

LT18: Change state to RDA, transmit acknowledgement, ACK.

6. Intercomputer Link Management Process LL

The intercomputer link management process LL implements another finite automaton (see Table 2). This automaton expresses the major content of the HL protocol.

The rows in the table indicate the possible states of circuit ports: sending data (SDA), receiving a pause request (RPS), receiving data (RDA), transmitting pause request (SPS), and initial link state (INIT).

Table 2. Automaton LL

Primary station

State / Event	SDA	RPS
RR	LL1	LL7
REJ	LL2	LL8
RNR	LL3	LL9
TMO	LL4	LL4
NNO	LL5	LL5
NNA	LL5	LL5
DIS	LL6	LL6

Secondary station

State / Event	RDA	SPS
I	LL10	LL12
YBF	LL9	LL13
NBF	LL11	LL9
DIS	LL6	LL6

Initial state

State / Event	INIT
SABM	LL14
UA	LL15

The columns in the table indicate the events that could appear: ready to receive (RR), reject (REJ), not ready to receive (RNR), beyond time limit (TMO), beyond the time limit N times in a row (NNO), rejection N times in a row (NNA), dismantle circuit (DIS), information frame (I), have buffer (YBF), no buffer (NBF), set up asynchronous balanced mode (SBA), and agree to set up asynchronous balanced mode (UA).

LLi in the table indicates the action that should be taken when the event occurs under the corresponding state.

LL1: Release buffer occupied by acknowledged information frame and transmit following information frame, clear beyond-the-time-limit count and reject count.

LL2: Retransmit current I frame, increment reject count by 1.

LL3: Change primary station state to RPS.

LL4: Retransmit current information frame and increment beyond-the-time-limit count by 1.

LL5: Notify operator that there have been N times no responses in a row or N times REJ in a row.

LL6: Change circuit state to initial state INIT and transmit DIS frame.

LL7: Change primary station state to SDA and transmit next information frame.

LL8: Change primary station state to SDA and transmit current information frame.

LL9: No action.

LL10: Check whether or not frame sequence number is right and if so, pass data received to packet level, increment reception variable by 1, transmit RR frame; if frame sequence number is wrong, discard information sequence received and transmit REJ frame.

LL11: Change secondary station state to SPS.

LL12: Transmit RNR frame.

LL13: Change secondary station state to RDA.

LL14: Set primary station state to SDA, set secondary station state to RDA and transmit UA frame.

LL15: Set primary station state to SDA, set secondary station state to RDA and transmit I frame.

7. Call-Response Process AC

The call-response process is a timing task. It manages the call bell signal and carrier frequency signal of 16 telephone circuits. The state of each telephone circuit changes as illustrated in Figure 5. After it is discovered that there is a tell signal on a circuit, the data assembly of one computer is set to connect and if the carrier frequency signal is received within the specified time (say, 2 minutes), then the circuit is initialized, its circuit characteristic table is set to the initial value so that the circuit can enter the communications state; if the carrier frequency signal is not received within the specified time, or if the carrier frequency signal is not constantly received in the communications process, then the circuit is set to idle and waits for the user to come and use it.

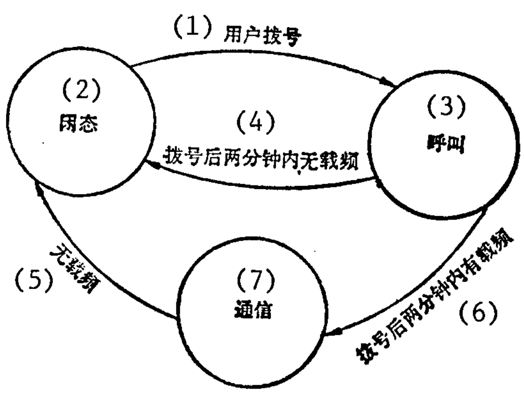


Figure 5. Diagram of Telephone Line State Change

Key:
1. User dial number
2. Closed state
3. Call
4. No carrier frequency within 2 minutes of dialup
5. No carrier frequency
6. Carrier frequency within 2 minutes of dialup
7. Communication

8. Distributed Data Process DD

The function of the distributed data process is to register the data which the packet level has passed to the link level in the transmission queue of the intercomputer circuits and if a circuit is in initialized state, the DD process enables the initialization of that circuit. The aim of designing this circuit is to provide a simple and accurate interface for the packet level.

IV. Interface Between Link Level and Other Parts

The link level operations under RDOS support and the interrupt module is the RDOS user self-defined facility interrupt processing program. Other processes are RDOS user processes. Actually, the link level relies very little on RDOS. Without RDOS support one only has to write a small interrupt general program and process call program to combine with link level software to support upper level network software. This is currently implemented on the SK1 computer.

The link level hardware interface is simple, and though link level processes four types of circuits, these four circuits are connected to DJS-100 series computers asynchronous multiplexers and can receive and transmit information in character interrupt mode.

There are a total of three interface commands between the link level and the packet level:

(1) PLDA (L-number, P-address, P-length)

(2) LPDA (L-number, P-address, P-length)

(3) LPER (L-number, Cause)

Here, L-number is line number, P-address is packet address, P-length is packet length, Cause is the reason for the trouble.

The packet level uses LPAD command to transmit the data between machines to the link level. The link level uses LPDA and LPER commands to pass the data received on line between machines as well as fault data received on line to the packet level.

Ther are only two interface commands beteen the link level and PAD:

(1) LPAD (L-number, D-address, D-length)

(2) PADL (L-number, D-address, D-length)

Where L-number the line number, D-address is the data block address, and D-length is the data block length.

The link level uses the LPAD command to pass the data received over the terminal circuit to PAD. The PAD process uses the PADL command to pass the data received over the terminal circuit to the link level.

V. Conclusion

Link level software was first developed 2 years ago when communication between machines still had to be by means of asynchronous interfaces. Its purpose was to implement synchronous duplex transparent transmissions under poor conditions and thus support software at the upper level of the network. Now, we are replacing it with synchronous interface and since we have adopted a modular structure, this change had not had a great influence on the overall link level software.

8226/6091
CSO: 4008/243

APPLIED SCIENCES

PACKET LEVEL PROTOCOL, ITS IMPLEMENTATION

Beijing JISUANJI XUEBAO [CHINESE JOURNAL OF COMPUTERS] in Chinese Vol 7, No 1, Jan 84 pp 20-30

[Article by Zhang Yingzhong [1728 7336 0022] of the Institute of Computing Technology, Chinese Academy of Sciences and He Rendong [6320 0088 2767] of the Institute of Information Control, Beijing; paper received 23 November 1982]

[Text] Abstract: The application of X.25 (1977) to packet-level protocol of RDC network was investigated. Problems concerning virtual circuit, packets and their confirmation, flow control, logical channel, state diagram and state table are first discussed, then packet-level module design, including data structure, commands of interface to both transport-level and link-level, routing algorithm, etc. A flowchart of packet-level implementation is also given.

I. Introduction

Packet level, also called network level, is situated in-between the link level and the transport level. On the basis of error-free transmission provided by the link level, it provides the transport level with data transmissions between two DTE. With the development of computer networking, CCITT issued proposal X.25 which described the interface rules between a DTE (data terminal equipment, such as host computer, smart terminal, and concentrator) and DCE (data circuit terminal device) on a public data network. After the X.25 proposal was issued, many computer networks internationally adopted X.25 as the basis of their network protocols and demanded that packet level protocol of RDC nets should observe the third level of X.25 (revised in 1977) to unify international standards.

This essay describes some concrete considerations when the X.25 proposal is applied to RDC Net packet level protocols, then describes some problems to be defined and resolved in designing the packet level modules, such as the design of interface commands between the transport level module and the link level module, design of data structures, and path algorithms. Finally, the essay introduces the general frame portion in the block diagram so that the reader will have an understanding of the packet level process.

II. Considerations in Applying the X.25 to RDC Net Packet Protocols

1. General Considerations

An RDC Net is a computer network made up of a number of computers connected together, the configuration can connect any number of computers into a network, but it does not indicate which computers at the nodes (called nodes or stations) are DTE and which are DCE, but it does require that:

(1) The same computer should be both host computer and communications node.

(2) Allow two computers as hosts to communicate directly.

For this reason, when designing the packet level module nodes they had to have both DTE functions and DCE functions; when the nodes are in DCE position they execute the operations which X.25 defines for DCE and when they are in DTE position, they execute the operations which X.25 defines for DTE. As how to define the state of a node situated in what position, we stipulate that:

(1) A terminal node which receives or sends data (on a virtual circuit, see section one on virtual circuits) is viewed as being in DTE position.

(2) If a local node is connected with a node in the DTE position (on a virtual circuit), then the local node appears in DCE status.

(3) When two nodes are both directly communicating as DTE then each node still must act as DCE in interfacing with the other party.

2. Virtual Circuits

The packet level functions lie in establishing a virtual circuit through calling between two DTE's that want to communicate and then carrying out data packet and interrupt packet transmission over the virtual circuit. To ensure that many virtual circuits can exist on a DTE/DCE interface at the same time a logic channel concept is used. The link between DTE/DCE is duplicated to make many logic channels, thus permitting each DTE to be able to link up with many DTE simultaneously and between each pair of DTE's many connections can be set up simultaneously and different virtual circuits can have the same actual path.

The link between two neighboring (i.e., having the same communications circuit) nodes on an RDC Net is reused as a multiple logic channel, one of which is used by each virtual circuit. Figure 2.1 is a diagram of three virtual circuits. The link between A/B, A/C, and B/D are reused as logic channels:

Virtual circuit (1) uses one logic channel between A/B and one logic channel between B/D, these two logic channels link up to form a virtual circuit. The terminal points A, D of this virtual circuit are in DTE position, but the node B is in DCE position.

Virtual circuit (2) uses one logic channel between A/C, A/B, B/D. These three logic channels link up to form one virtual circuit. The terminal points C and D of this virtual circuit are both in DTE position, but nodes A and B are in DCE position relative to C and D respectively.

Virtual circuit (3) uses one logic channel between B/D. This logic channel makes up a virtual circuit. At this time B and D are in DTE position but each is also in DCE position relative to the other.

In the B/D interface in Figure 2.1 there are three virtual circuit passages, two on A/B interface and one on A/C interface.

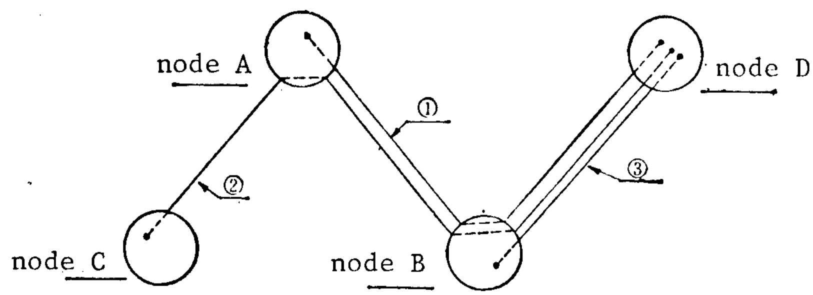


Figure 2.1. Virtual Circuits and Logic Channels of an RDC Net

3. Packet Level and Its Response Relationship

It is packets that are transmitted on logic channels. Some are used for transmitting control information and some are used for transmitting data. Same with X.25 a total of 14 packets have been designed for the RDC Net. To simplify discussion, we have listed these packets and given them abbreviated names and shorthand symbols (see Table 2.1).

Table 2.1. Packet Symbols, Abbreviations and Shorthand Symbols

Packet symbol	Abbreviation	Symbol	Functions
00001011	Call request	CAR	Requests setting up a virtual circuit with the DTE called on the designated channel.
00001111	Call confirm	CAC	Confirms request to set up a virtual circuit with the DTE called on the designated circuit.
00010011	Clear request	CLR	Requests clearing virtual circuit on the designated logic channel and giving reasons for clearing.
00010111	Clear confirm	CLC	Confirms request for clearing virtual circuit on designated logic channel.

[Table continued on following page]

Packet symbol	Abbreviation	Symbol	Functions
P(S)OP(R)O	Data packet	DATA	Transmitting data packet designated by P(S) to the other party over the designated logic channel and notifying the other party that this party is ready to receive data packet designated by P(R) over the logic channel.
00100011	Interrupt request	INR	Transmits interrupt packet over the designated logic channel and gives interrupt information.
00100111	Interrupt confirm	INC	Confirms interrupt request sent by other party on the designated channel.
P(R)00001	RR	RR	Notifies other party that this party is ready to receive data packet designated as P(R) on the designated logic chennel.
R(R)00101	RNR	RNR	Notifies other party that this party cannot continue to receive further data packets on the designated logic channel.
P(R)01011	REJ	REJ	Notifies other party of request to retransmit data packet designated P(R) and subsequent data packets.
00011011	Reset request	SER	Requests that virtual circuit be reset on designated logic channel and gives reason for reset.
00011111	Reset confirm	SEC	Confirms request to reset virtual circuit on designated logic channel.
11111011	Reactivate request	STR	Used to simultaneously clear all called virtual circuits on the interface between two nodes and resets all permanent virtual circuits.
11111111	Reactivate confirm	STC	Confirm reactivate request.

Note: In the table the logic channel is the logic channel designated by the logic channel number and logic channel group number in the packet. This party refers to the terminal node which is transmitting the data packet or controlling the packet and the other party refers to the terminal node which is receiving the data packet or controlling the packet, P(S) and P(R) both take up three bits and take values from 0 to 7.

When the RDC packet level implements the above 14 packet receptions and transmissions it uses the following methods:

(1) Since setting up a virtual circuit is something which has overall significance, end-to-end confirmation mode is used for call requests. The packet level is responsible for transmitting the CAR packet by begin with calling DTE node by node to the DTE called by a certain path algorithm. On the actual path used, between each station the virtual circuit linked by the logic channel used by the CAR packet is transmitted and when the DTE called agrees to the call request, the packet level transmits the CAC packet back to the DTE called along the same virtual circuit and after this, the virtual circuit enters the duplex data transmission stage.

In the RDC Net, a virtual circuit established in the calling stage cannot be changed until the virtual circuit is cleared.

(2) Point-to-point confirmation mode is used for clearing requests because clearing is a forced command, especially when there is a rule error, the virtual circuit can no longer transmit data thus it is best to release the logic channel resources occupied by the virtual circuit as quickly as possible.

(3) End-to-end confirmation mode is used for data packet and interrupt packet. Data packet end-to-end confirmation means that only the DTE which is receiving the data packet can give the corresponding data packet receiving sequence P(R). Other stations only play a role in transmitting P(R). And interrupt packet end-to-end confirmation means that only a DTE which receives an interrupt packet can give interrupt confirmation.

(4) Point-to-point confirmation mode is used for the reset request. Generally speaking point-to-point confirmation mode can make packet level processing simple and highly efficient and it can also be unified with the clear request processing mode. However, unlike the clear request, the aim of reset is to initialize the virtual circuit and not to release logic channel resources.

(5) End-to-end confirmation mode is used for restart request.

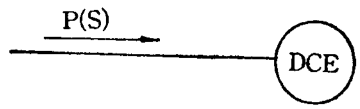
(6) In implementing the packet level, since memory resources are limited, the packet level modules should not be too big, therefore we imposed limitations on optional user tasks and unified the window size and the maximum packet length, but we have not yet established a time limit mechanism.

4. Flow Control Structure

To keep the data packets from unlimited crowding of the packet exchange network, X.25 introduced window structure for control of the data packets on each virtual circuit.

RDC Net flow control is implemented separately at both ends of the DTE/DCE interface and since it uses uniform maximum packet length and uniform window size, RDC Net flow control can be designed as follows:

First, at the DCE node



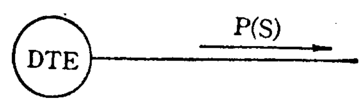
carries out a window check on the P(S) received (i.e., the data packet received with the transmission sequence number P(S), similar situations will not be explained again), to see whether it satisfies

$$WPR \leqslant P(S) < WPR + WS$$

(in which WPR is the value of the P(R) most recently sent by the station, i.e., the value of the P(R) carried in the RR packet and the data packet sent to the DTE; WS is the window size, ditto below). If it is satisfied, then it transmits this data packet, otherwise it initiates reset.

Next, at the DTE node

(1)

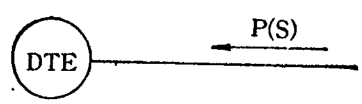


carries out a window check on the P(S) transmitted to see if it satisfies

$$LPR \leqslant P(S) < LPR + WS$$

(in which LPR is the value of P(R) most recently received by this station). If it is in the window, then this data packet is sent, otherwise it is not.

(2)

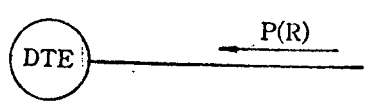


A sequential check is carried out on the P(S) received to see if it satisfies

$$P(S) = IPR$$

(in which IPR is the value of the P(R) most recently sent by this station, IPR = the value of the last P(S) received + 1). If it is not satisfied, then it initiates reset. Those which satisfy the sequential check definitely satisfy the window check.

(3)



A range check is carried out on the P(R) received to see if it satisfies

$$LPR \leqslant P(R) \leqslant PS$$

(in which PS is the value of the P(S) of the data packet which this station is ready to send). If it is satisfied, the P(R) is made the lower edge of the window, otherwise, reset is initiated.

5. State Diagram of the Logic Channels

Packet reception and transmission is carried out on certain logic channels. For systematic work, X.25 provides logic channel state diagrams. When carrying out packet level rules, changes in the logic channel state are considered at the station on one side of the logic channel node.

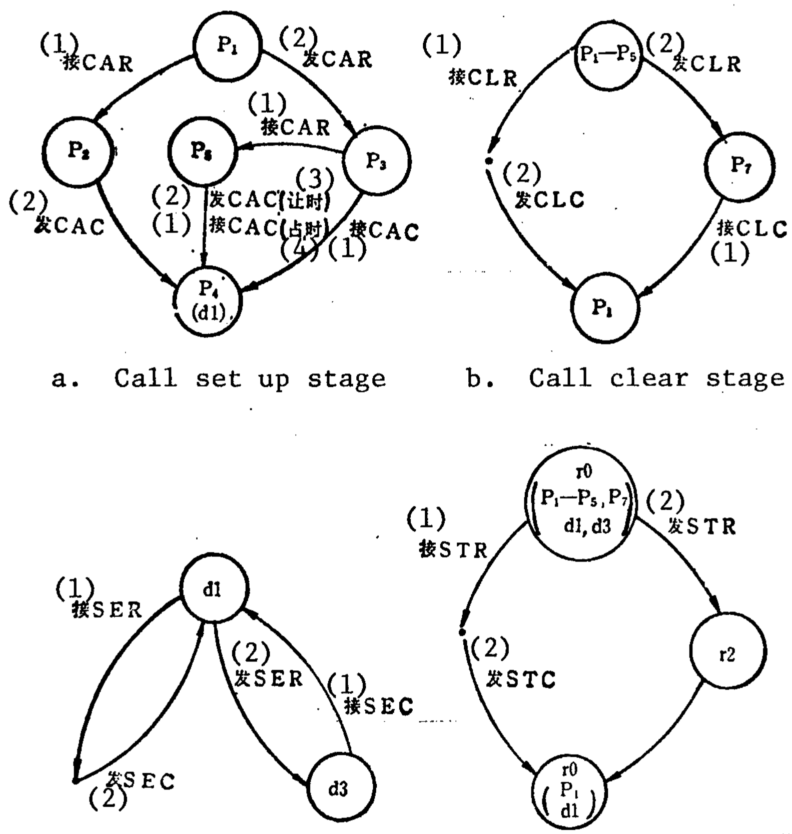
In the light of the call set up stage, X.25 stipulates that when a call collision occurs, the DCE should let the call request through and cancel the call directive. This point is reflected in that in the implementation of RDC Net packet level in two neighboring nodes one sets a "yield" marker and the other sets an "occupied" flag in the physical link between them (see section one part four) and when a call collision occurs, the node marked "yield" will process according to the actions stipulated for DCE in X.25.

In view of the stipulations on packet response relations in section three part two, for example, that the point-to-point confirmation mode is adopted for clear requests, then P_6 state is only implied and is not repeated. Similar situations can be analogized for reset and restart. Again, in view of the fact that in packet level implementation a time limit mechanism has not yet been set up, thus we derive the practical logic channel state diagram illustrated in Figure 2.2.

Since restart rules are used in clearing all calling virtual circuits on the DTE/DCE interface and resetting all permanent virtual circuits, therefore, what r0 and r2 in the diagram each actually refer to is the state of the physical link between interfaces.

6. State Matrix Tables

In addition to logic channel state diagrams, X.25 also formulates for the DCE the action format which should be adopted when the DTE accepts various packets under various states. Generally speaking, this also can be used as reference for the DTE. Combining the action tables and the logic channel state diagrams and considering them one can derive a state matrix table for writing the packet level module. This form is very visual and is not hard to derive so, due to space limitations, we will not deal with it here.



a. Call set up stage b. Call clear stage

c. Reset stage d. Restart stage

Figure 2.2. Logic Channel State Diagram

Key:

1. Receive
2. Send
3. When yield
4. When occupied

III. Design of Interface Commands Between Neighboring Second Level Modules

1. Interface Relations

In a node, the packet level module will interface with the transport level module TS and the link level module LL. The interface relations are illustrated in Figure 3.1.

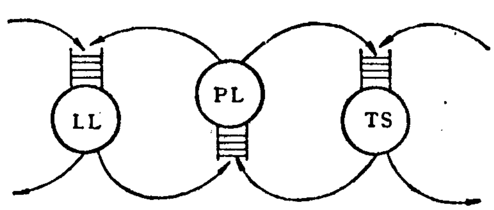


Figure 3.1. Interface Relations Between PL and TS and Between PL and LL

In an RDC Net, the communications between modules at the various levels is implemented by means of the two queue operations primitives PUT and GETS (and GET). Only one process was set up for the packet level (under present circumstances where three 100-series computers are linked together, it is a single task under RDOS control), thus PL also is used to express the packet level process. In an RDC Net, when any process needs the services of another process, through PUT it registers the processing unit in the relevant processing queue; through GETS (or GET), any one process accesses a processing unit from this processing queue and carries out the process demanded and thus serves another process. The processing unit format differs as the needs of the level differ.

The design of the interface commands is also the design of processing units. The processing unit between TS and PL is determined by what functions PL can provide and what demands TS has. The relation between PL and TS are packet level reception and transmission, therefore they are rather simple.

2. Interface Commands Between PL and TS

TS → PL:

TCAR--demands set up of a virtual circuit.

TIND--agrees to the request of a remote TS to set up virtual circuit.

TNCA--disagrees with request of a remote TS to set up virtual circuit.

TDATA--demands transmission of a data packet.

TINR--demands transmission of an interrupt packet.

TSER--demands reset of virtual circuit.

TCLR--demands clearing of virtual circuit.

PL → TS:

PCAR--relays to TS the request of a remote TS to set up virtual circuit.

PIND--reports to TS that remote TS has agreed to request to set up of a virtual circuit here.

PDATA--relays data packet to TS.

PINR--relays interrupt packet to TS.

PSER--reports to TS, virtual circuit is reset.

PCLR--reports to TS, virtual circuit is cleared.

3. Interface Commands Between PL and LL

PL → LL:

PLDA--demands LL send a packet on a designated circuit.

LL → PL:

LPDA--LL delivers a packet.

LPER--LL reports link trouble.

LPRC--actually, this is a controller command. By the controller command the analyzer program converts into the processing unit and notifies PL that the link with the problem has been repaired.

IV. Physical Link Tables and Logic Channel Tables

1. Physical Link Tables

In the RDC Net there are nodes connected by communications circuits which are called neighboring nodes. The packet level feels that there is a physical relationship between neighboring nodes. A node can have several physical links connecting it with other nodes.

To depict graphically the characteristics of physical links and the states in which they can exist, a physical link table should be set up. The table takes up three words and records:

<table>
<tr><td>0</td><td>1</td><td>… 4 …</td><td>6 7</td><td>…</td><td>10 … 15</td></tr>
<tr><td>P</td><td>U</td><td>0 0 0 0</td><td>r state</td><td>0 0</td><td>communications circuit number</td></tr>
<tr><td>0</td><td>0</td><td>0 0</td><td colspan="3">reused logic channel numbers</td></tr>
<tr><td>0</td><td colspan="5">logic channel table address of zero-th logic channel</td></tr>
</table>

in which:

(1) P bit is "yield" and "occupied" marker bit if

P = { 0, for yield: it indicates that the logical channel should be divided from small to large, and that when the call collision occurred, the station was processing in DCE position.
1, for occupied: situation just the opposite of the above.

(2) U bit is the trouble bit if

$$U = \begin{cases} 1, & \text{link trouble, LL reports final position of link trouble,} \\ & \text{controller commands indicate that it is repaired or to} \\ & \text{clear after other party issues a STR packet.} \\ 0, & \text{link is normal.} \end{cases}$$

(3) r state is restart stage state and occupies bits 6 and 7.

$$6,\ 7 = \begin{cases} 00, & \text{is r0 state.} \\ 10, & \text{is r2 state.} \end{cases}$$

(4) Communications circuit number is the QTY circuit number (one of $0\text{-}64_{10}$ lines) occupied on this node for the physical link.

(5) Reused logic channel number refers to how many logic channels are reused on this physical link. These logic channels are numbered sequentially beginning with 0.

(6) Logic channel table address of the zero-th logic channel, since logic channels are numbered sequentially beginning with 0, and the logic channel table is continuously arranged, the logic channel table address can be found on the basis of the logic channel table address and the logic channel number based on the zero-th logic channel.

2. Logic Channel Table

This table records necessary information about the logic channel and takes up 6 words:

0	1 ... 4 ...	7	8	...	10 11	...	13 ... 15
PVC		INT	FL		d state		p state
CHAIN (PHLI)				(LGID)			
	WS			WPR			
	PS			LPR			
	IPR			NUMBER			
	HEAD						

in which:

(1) PVC bit is the marker bit of the permanent virtual circuit.

$$PVC = \begin{cases} 0, \text{ this logic channel is not a permanent virtual circuit.} \\ 1, \text{ this logic channel is a permanent virtual circuit.} \end{cases}$$

(2) INT bit is the marker bit for whether or not an INR packet can be received again, only with an FL = 1 logic channel can the INT bit be filled.

$$INT = \begin{cases} 0, \text{ indicates that it can receive (or send) INR packets.} \\ 1, \text{ indicates that it cannot receive INR packets.} \end{cases}$$

(3) FL is flow control marker and indicates whether or not this station is in DCE position in flow control processing.

$$FL = \begin{cases} 0, \text{ indicates that this station is not in DCE position.} \\ 1, \text{ indicates that this station is in DCE position.} \end{cases}$$

(4) p state and d state occupy 3rd and 2nd bits respectively and are used to record the p state and d state of this logic channel. For example, if the value of the 3rd bit which is occupied by the p state is 1, it means that it is in p_1 state and the other states are similarly expressed.

(5) CHAIN is the chain syllable. The first four bits are called the PHLI syllable and the last 12 are called the LGID syllable. They are used to record which physical link (indicated by the first four bits) and which logic channel (indicated by the last 12 bits) this logic channel is chained with.

If the PHLI syllable = 0000, then this logic channel is connected with TS and the value of LGID at this time is the INDEX number used by TS to call this virtual circuit.

(6) WS syllable records the window measurements. Since P(S) and P(R) can be modulus 128, 7 bits are reserved, but currently modulus is 8, so it takes up only 3 bits. It is the same with WPR, PS, LPR, and IPR described below. For ways of using each syllable, see section 4 in part two.

(7) WPR syllable is used by FL = 1 logic channel to record whether the window and its value is equal to the P(R) value in the confirmation packet used (i.e., the data packet and RR packet).

(8) PS syllable records the P(S) value of the DATA packet issued when the station is a terminal node.

(9) LPR syllable records the P(R) value in the most recent confirmation packet received from the other terminal when the node is a terminal node.

(10) IPR syllable records the P(S) value of the DATA packet expected from the other terminal when the station is a terminal node.

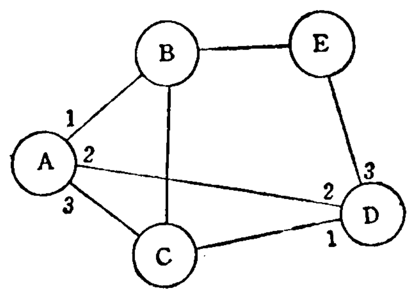
(11) NUMBER syllable refers to the data packet unit number waiting to be sent because of limitations in the window of this terminal.

(12) HEAD syllable records the head of the address chain of the data packet waiting to be sent.

Please note that both the physical link table and the logic channel table refer to the table recorded for the station at one end of that physical link and that logic channel, and of course, the node on the other side also has corresponding tables. They record information from the angle of each station and cannot have uniform records which override the station.

V. Path Algorithm

RDC Net currently uses a (selected) fixed path algorithm, i.e., before the fact, a path table is arranged for each node. Figure 5.1 gives the path table for node A when an RDC Net has 5 nodes. Of course, as the situation changes, the path tables also can be revised.



Node A path table

Target node	Second choice	First choice
A	0	0
B	3	1
C	1	3
D	3	2
E	2	1

Node D path table

Target node	Second choice	First choice
A	1	2
B	2	3
C	2	1
D	0	0
E	1	3

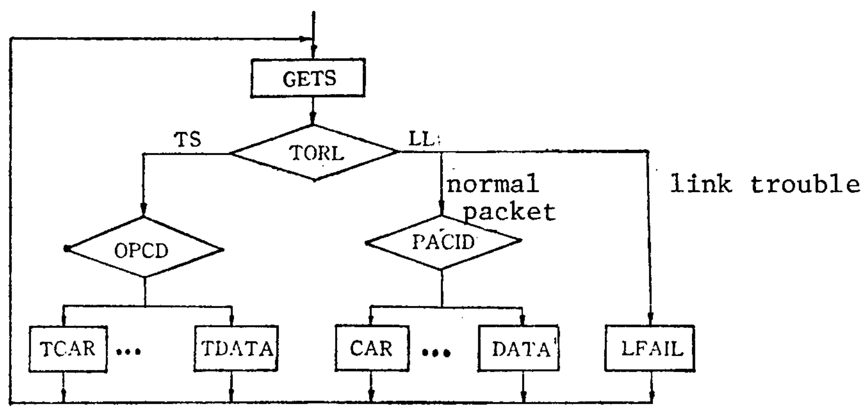
Figure 5.1. RDC Net Path Algorithm

For example, there is a call request packet at node A (only call request packets carry out path selection in an RDC Net) to be sent to target node E. It first of all selects link 1 to transmit the packet, and if problems occur on link 1 or there is no logic channel ready, then it transmits the packet on its second choice, link 2. When it transmits along the first choice, node B takes the packet and then transfers the call packet over node B path table; when node D receives the packet by the second choice transmission, D node first considers transmitting the packet on the logic channel on link 3 which is ready, then considers transmitting the packet on the logic channel on link 1 which is ready.

The nodes on the RDC Net now use an all-network unified numbering which will be made the network address of the node. Thus we can find the path information by the offset variable in numbering. Note that the link number of the nodes is not exactly 0, the 0 in the path tables indicates that the station is a target station.

VI. Introduction to the General Framework

The packet level process PL operates under the control of the general framework (see figure below), since the general framework constantly draws processing units from the processing unit queue for analysis and processing, each time a processing unit is drawn, there is a need to determine whether the processing unit is from TS or LL before it is processed.



in which the meaning of each frame is as follows:

(1) GETS: PL process draws a processing unit through GETS sequence.

(2) TORL: Determines whether the processing unit comes from TS or LL and if it is from LL then it must be determined whether it is a normal packet which has been passed or a report of link trouble and then process it in PACID or LFAIL frames.

(3) OPCD: When a processing unit is passed from TS, the processing unit should be analyzed to determine which kind of the TS → PL it is and transfer it to the sub-frame. Since there are seven different processing units, there are seven sub-frames which use names similar to the processing units: TCAR, TIND, TNCA, TDATA, TINR, TSER, TCLR. TCAR processes requests to set up virtual circuits; TDATA processes requests to transmit data packets; ...

(4) PACID: Consists of public information, and analyzes which type of packet is passed from LL, i.e., transfer it to the sub-frame depending on the packet marker. Since there are 14 packets, there are 14 sub-frames, which use abbreviated names similar to those of packets: CAR, CAC, CLR, CLC, DATA, INR, INC, RR, RNR, REJ, SER, SEC, STR, STC. CAR processes call request packets received; DATA processes data packets received; ...

(5) FAIL: Carries out appropriate work when there is link trouble.

Thus, the packet level modules are made up of the general framework and 21 sub-frames. After the processing of each frame and the LFAIL frame is completed they return to the general frame GETS, and continue to draw processing units and analyze processing until finished.

VII. Conclusion

The RDC test net has already gone into operation and has satisfied the demands of users transmitting data over the network. The packet level modules we have designed provide some experience for future implementations on packet level protocols. Currently, the packet level modules have only basic functions and need to be strengthened and improved for optional user tasks, time limit mechanisms, and path algorithms as the RDC Net moves into a new stage. In February 1980 CCITT issued revised X.25 proposals which increased such important service functions as digital data reporting and the RDC Net packet level also should give this full consideration.

8226/6091
CSO: 4008/243

LIFE SCIENCES

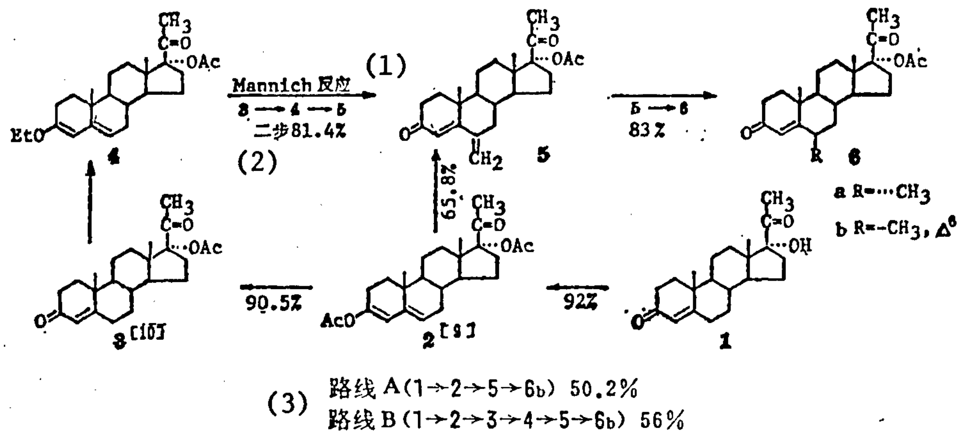
NEW SYNTHETIC ROUTE FOR MEGESTROL ACETATE REPORTED

Shanghai YOUJI HUAXUE [ORGANIC CHEMISTRY] in Chinese No 6, Dec 83 pp 451-453

[Article by Ni Yuan [0242 0337], Yu Xindi [0151 2450 1229] and Cai Zuhui [5591 4436 1926] of the Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, and Zhang Shuiquan [1728 3055 3123] of the Shanghai Pharmaceutical Plant No 12: "A Simple Synthetic Route for Δ^6-6-Methyl-17a-acetoxy-progesterone (Megestrol Acetate)"; Received 29 November]

[Text] Megestrol acetate (Δ^6-6-methyl-17a-acetoxy-progesterone, 6b) is an effective steroid contraceptive.[1,2]

The steroid is generally introduced in the following three ways: By applying the Mannich reaction to 5,6-epoxide produced through peracid reaction;[3,4] by producing dibromomethylene via carbon tetrabromide-pyridine reaction and then hydrogenizing and isomerizing the product;[5] or by producing C_6-methylene (5) through the Vilsmeier method and isomerizing this product with palladiumcarbon.[6] The latter method is currently employed in China. Recently, Annen et al. have reported a simplified method of producing the C_6-methylene steroid (5).[7] This method involves a single step in which 17a-acetoxy-progesterone (3) is treated with diethylmethane (or dimethoxymethane) in a phosphorus oxychloride reaction. Our laboratory employed $\Delta^{3,5}$-3,17a-diacetoxy-progesterone (2) and the reaction conditions reported by Annen et al. to produce C_6-methylene (5) in a yield of 65.8 percent. Since $\Delta^{3,5}$-3,17a-diacetoxy-progesterone (2) is a precursor of 17a-acetoxy-progesterone (3), the synthetic procedure can be shortened. We also employed enol ether (4) from 17a-acetoxy-progesterone (3) to obtain C_6-methylene (5) via the Mannich reaction and then isomerized this product with palladium/calcium carbonate to obtain our desired product, megestrol acetate (6b). As the diagram indicates, route A ($1 \rightarrow 2 \rightarrow 5 \rightarrow 6b$) produces megestrol acetate (6b) in a total yield of 50 percent, which is equal to the highest yield reported in the literature, but the process is shortened. Route B ($1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6b$) produces a total yield of 56 percent, which is better than the highest yield reported in the literature.[6]



Key: 1. Reaction 2. Two steps 3. Route

Generally, methyl progesterone (6a) and megestrol acetate (6b) are both prepared by isomerizing C_6-methylene (5) with a palladium/hydrocarbon in the presence of cyclohexene. We achieved a high yield of megestrol acetate (6b) by isomerizing C_6-methylene (5) with palladium/calcium carbonate in the presence of sodium acetate and one percent cyclohexene. Even though the reaction time was prolonged approximately 16-fold, the methyl progesterone 240 nm absorption peak did not appear in the reactants under ultraviolet spectrometry.

Experimental

Although demonstrated, the melting point was not corrected. We employed the WZZ-1 to compare polarization, the Shimadru UV-200 and the UV-730 for ultraviolet spectrometry, the UR-10 for infrared spectrometry, the Varian XL-200 and the EM-360 to determine NMR and the JMS-01V for mass spectrography.

6-Methylene-17a-acetoxy-progesterone (5)[6a]

(A) In a 250 ml ground glass flask, we added in succession 1 g of $\Delta^{3,5}$-3,17a-diacetoxy-progesterone (2), 1 g sodium acetate, 30 ml chloroform, 30 ml diethylmethane and 3.8 ml phosphorus oxychloride and reacted and treated this mixture according to the conditions reported in the literature.[7b] We employed dichloromethane-methanol to recrystallize the residue, obtaining 400 mg of C_6-methylene (5), m.p. 227-229°C, and recrystallized the solution again, m.p. 234-236°C, obtaining an additional 210 mg of C_6-methylene (5) for a total yield of 65.8 percent, $[a]_D^{20}$ + 197°($CHCL_3$, C, 0.85).

UV λ_{max}^{EtOH}: 260nm(ε 16800);

IR ν_{max}^{nujol}: 3075 (C_6-CH_2) 1735、1260 (C_{17a}-$OCOCH_3$), 1610、1675(Δ^4-3-keytone), 1710 (C_{20}-keytone)cm^{-1}。

^{1}H NMR δ value ($CHCl_3$,60 MHz,TMS):0.65 (*s*,3H, C_{18}-CH_3),1.09(*s*,3H,C_{19}-CH_3), 2.06 (*s*,3H,C_{17a}-$OCOCH_3$),2.12(*s*, 3H, C_{21}-CH_3), 5.0、5.10(2*m*,$W_{1/2}$ = 4Hz, 2H,C_6-CH_2), 5.98 (*s*,1H,C_4-H)。

MS *m/z*:384(M^+)。

(B) In a 100 ml round-bottom flask, we added 5 g of 17a-diacetoxy-progesterone (3), 50 ml absolute benzene, 7 ml absolute alcohol, 3.9 ml triethylorthoformate and 21 mg paratoluene sulfonic acid and heated, stirred and refluxed this mixture in an oil bath for 70 minutes. When the solution cooled to room temperature, we added 1 ml pyridine, and after the solution concentrated under reduced pressure, we added 50 ml ethyl alcohol and 25 ml chloroform, which we dissolved by heating the mixture. When the latter cooled to 10°C, we added 1.9 ml methylaniline, 7.5 ml of 37 percent formaldehyde and 120 mg paratoluene sulfonic acid; stirred the mixture for 20 minutes; and then reacted the ingredients at 60°C for 1 hour. Next, we added 2 ml pyridine, and after the solution had concentrated under reduced pressure, we added 20 ml methanol and shook the mixture. When the latter cooled to 5°C, we added 20 ml concentrated hydrochloride and stirred the mixture for another 2.5 hours. Then we removed solids through filtration, neutralized the filter cake by washing it in water and dried the product, obtaining 4.88 g of a yellow powder, which we recrystallized by adding ethyl acetate, obtaining 4.2 g of a darkish-white crystal, m.p. 232-235°C, in a yield of 81.4 percent. The $[a]_D$, IR, NMR, MS produced were exactly the same as the C_6-methylene (5) obtained in method A.

Δ^6-6-Methyl-acetone-progesterone (6b, megestrol acetate, megestrol acetate ester acetate)

In a 500 ml round-bottomed flask, we added 10g C_6-methylene (5), 200 ml absolute alcohol, 4 g anhydrous sodium acetate, 7 g palladium/calcium carbonate and 1 percent cyclohexene and heated and stirred this mixture for 1 hour. We employed ulitraviolet spectrometry, and when the characteristic absorption peak of only 288 nm for megestrol acetate (6b) was attained, we terminated the reaction. While the solution was still hot, we filtered it, concentrated the filtrate under reduced pressure, dried this product, added cold water, crushed the resulting solids, filtered the solution again and neutralized it. When the mixture dried, 9.7 g of a yellow powder was obtained, which we recrystallized with dichloromethane ethylate, obtaining 8.3 g of a white crystal in a yield of 83 percent, m.p. 217-219°C, $[a]_D^{17}$ + 11.5°($CHCL_3$).

UVλ_{max}^{EtOH}: 287 nm(ε 25400)。

IR ν_{max}^{nujol}: 1730、1250($C_{17\alpha}$-$OCOCH_3$), 1700 (C_{20}-ketone), 1580、1620、1660($\Delta^{4,6}$-3-ketone) cm^{-1}。

^{1}H NMR δ value ($CHCl_3$, 60MHz, TMS): 0.70 (*s*, 3H, C_{18}-CH_3), 1.10(*s*, 3H, C_{19}-CH_3), 1.89 (*s*, 3H, C_6-CH_3)、 2.00(*s*, 3H, $C_{17\alpha}$-$OCOCH_3$), 2.06(*s*, 3H, C_{21}-CH_3), 5.77(*s*, 1H, C_4-H), 5.9 (*s*, 1H, C_7-H)。

MS *m/z*: 384(M^+)。

REFERENCES

[1] a. Nagata Wataru and Itazaki Hiroshi, in "Shin Iyakuhin no Goseiho" [Methods for Synthesis of New Drugs], Vol 2, ed. Tsuda Kyosuke and Nogami Hisashi, Tokyo, Chijin Shokan, 1972, p 525.

b. V. Petrow, CHEMICAL TECHNOLOGY, No 7, 1977, p 563.

[2] a. Shanghai Institute of Organic Chemistry, Department 1, Group 101, YOUJI HUAXUE [ORGANIC CHEMISTRY], 1981, p 149.

b. Yang Yiqian [2799 0110 6197] et al., SHENGZHI YU BIYUN [REPRODUCTION and CONTRACEPTION], Vol 2, No 4, 1982, p 23.

[3] H. J. Ringold et al., JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, No 81, 1959, p 3712.

[4] J. C. Babcock et al., Ibid., No 80, 1958, p 2904.

[5] S. Liisberg et al., TETRAHEDRON, No 9, 1960, p 149.

[6] a. D. Burn et al., Ibid., No 20, 1964, p 579.

b. D. Burn et al., Ibid., No 21, 1965, p 1619.

[7] a. K. Annen et al., Ger. offen. DE 3,004,508 (1981); CHEMICAL ABSTRACTS, No 96, 1982, p 6951.

b. Idem., SYNTHESIS, 1982, p 34.

[8] a. Van Rheenan and H. Verlan, Ger. Offen. DE 2,055,895 (1971); U.S. 3,642,840 (1972); CHEMICAL ABSTRACTS, No 75, 1975, p 49443j.

b. Cai Zuhui et al., YAOXUE XUEBAO [ACTA PHARMACEUTICAL SINICA], No 17, 1982, p 233.

[9] N. R. Pemmaraju and B. E. Edwards, U.S. 3,321,495 (1967); CHEMICAL ABSTRACTS, No 67, 1967, p 117103k.

[10] M. Windholz, "Merck Index," 9th ed., New Jersey, Merck Co, 1976, p 4756.

12431/9604
CSO: 4008/231

END